NORTHWESTERN UNIVERSITY

Essays in Labor Economics

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Basit Zafar

EVANSTON, ILLINOIS

June 2008

Table of Contents

List of Tables

List of Figures

ABSTRACT

Essays in Labor Economics

Basit Zafar

This dissertation analyzes how individuals choose college majors. The choice of college major is treated as one made under uncertainty. Understanding any decision under uncertainty requires one to study how expectations and preferences are used to make the choice. However, since observed choices may be consistent with many combinations of expectations and preferences, I instead collect a unique panel dataset of Northwestern students which contains their subjective expectations about choice-specific outcomes.

Chapter 2 estimates the decision rule of college major choice by combining subjective expectations with choice data. I obtain three main results: (1) non-pecuniary outcomes explain nearly half of the choice, (2) males and females are similar in their preferences for outcomes in college but differ in their preferences for outcomes in the workplace, and (3) the gender gap in major choice is mainly because of gender differences in beliefs about enjoying studying different majors, and gender differences in preferences.

Chapter 3, motivated by the fact that there is a positive correlation between one's own major and that of their parents and elder siblings, outlines a model in which conformity

in actions may arise from learning about the norm, or from image-related concerns (social influence). To empirically disentangle the two, I use the fact that image-related concerns can only be present if actions are publicly observable. The model predictions are tested in a charitable contribution experiment in which the actions and identities of the subjects are unmasked in a controlled and systematic way. Both learning and social influence seem to play an important role in the choices of the subjects.

Chapter 4 focuses on how individuals revise expectations, and analyzes perceptions of discrimination associated with major choice. Changes in expectations are found to vary in sensible ways. Priors for outcomes realized in college are found to be fairly precise, while students seem to gain valuable information about outcomes that are realized in the workplace. Perceptions of being treated poorly in the jobs in the various majors are found to be negatively correlated with the fraction of one's own gender in that field of study.

Acknowledgements

I would like to thank members of my dissertation committee, Charles Manski, Paola Sapienza, and Christopher Taber for their invaluable guidance and continuous support. I am particularly grateful to Charles Manski for being an outstanding advisor. I am also grateful to Raquel Bernal, Hilarie Lieb, and participants of the Northwestern Labor Lunches for their insightful comments and suggestions. I would like to thank all those involved in providing the datasets used in this dissertation. I also want to thank all the people who made my PhD years in Evanston a joyful adventure, and in particular Joyee Deb, Aniruddha Gupta, Piyush Jain, Siddarth Madhav, Zahra Siddique, and Amir Toft. Financial support for my dissertation was provided in part by the Graduate Research Grant, Eisner Fund, and Ronald Braeutigam.

Above all, I wish to thank my parents, Nargis Perveen and Zafar Iqbal, for their constant care, love and encouragement over the years. They've always been there for me when I needed them, and have been fully supportive of my choices. Finally, I wish to thank my wife, Samina. She followed me on this voyage, lived through its ups and downs, and kept me afloat through my hard times. I dedicate this dissertation to my family- I feel they share a large part of my achievements and that I could not have done it alone.

To my family

CHAPTER 1

Introduction

Choosing a college major is a decision that has significant social and economic consequences. However, little is known about how youth choose college majors. A second intriguing point in the context of college majors is the empirical fact that males and females choose very different majors. For example, in 1999-2000 in the United States, while nearly three-quarters of the recipients of Education bachelor's degrees were females, less than one-fifth of Engineering bachelor's degree recipients were females (Dey and Hill, 2007). The first part of this dissertation focuses on the question of how undergraduates choose college majors, and attempts to explain why males and females make different choices with regards to college majors.

I treat the choice of college major as one made under uncertainty- uncertainty about personal tastes, individual abilities, and realization of outcomes related to choice of major. Understanding any decision under uncertainty requires one to study how expectations and preferences are used to make the choice. The approach prevalent in the literature is to make non-verifiable assumptions on expectations, and employ choice data to infer preferences. However, this can be problematic since observed choices may be consistent with many combinations of expectations and preferences (Manski, 1993a). In order to overcome this identification problem, I collect additional data on expectations to estimate a choice model of college majors. The study was conducted at Northwestern in Fall 2006 and Fall 2007; the dataset contains students' subjective expectations about choice-specific outcomes, and data on their demographics and background information.

In Chapter 2, I estimate a random utility model of college major choice allowing for heterogeneity in beliefs and controlling for both the pecuniary and non-pecunairy determinants of the choice. Prior to this work, there has been virtually no empirical analysis of the non-pecuniary determinants of the choice of college majors. This gap in the literature primarily stems from a lack of detailed data on the non-pecuniary outcomes of the choice. I find that non-pecuniary outcomes are significant in the choice. Enjoying coursework, enjoying work at potential jobs, and approval of parents are the most important determinants in the choice of college major. Males and females have similar preferences while in college, but differ in their preferences in the workplace; males care more about pecuniary aspects (social status of the jobs, future income) while females care more about the non-pecunairy aspects of the workplace (enjoying working at the jobs, reconciling work and family). The second half of chapter 2 focuses on the underlying reasons for the gender gap in the choice of majors. At least two different explanations have been put forward in the literature for this gender gap: (1) innate differences between males and females (Kimura, 1999; Baron-Cohen, 2003), and (2) gender-based discrimination (Valian, 1998). The structural approach that I adopt in the paper allows me to check the validity of these hypotheses. I decompose the gender gap into differences in beliefs and preferences. First, I find that gender differences in beliefs about academic ability and expected income constitute a small and insignificant part of the gap; this allows me to rule out hypotheses like women being low in self-confidence relative to men (Niederle et al., 2007), and monetary discrimination in the workplace as possible explanations for the gender gap. Conversely, I find that most of the gender gap is due to differences in beliefs about enjoying coursework, and gender differences in preferences.

Chapter 2 does not focus on the aspect that individuals may find it optimal to experiment with different majors to learn about their ability and match quality (Manski, 1989; Altonji, 1993; Malamud, 2006). This is one of the issues explored in Chapter 4. More specifically, this chapter tries to address three questions: (1) why and how individuals revise their expectations for the various major-specific outcomes, (2) why females, relative to males, enjoy studying fields like engineering and sciences less, and (2) why individuals experiment with different majors. For this purpose, the students who were surveyed for chapter 2 were re-surveyed. I find that changes in expectations about various major-specific outcomes vary in sensible ways. Moreover, priors for outcomes that are realized in college (like approval of parents, graduating in 4 years) are fairly precise, while individuals seem to gain valuable information between the two surveys about outcomes that are realized in the workplace. Though individuals seem to be aware of a wage gap in favor of males in most majors, they underestimate the extent of the gap, and incorrectly believe that the wage gap stays roughly constant over time. Moreover, males and females differ in their reasons for the wage gap- while males believe it to be because of innate differences between the two genders, females believe it is because employers expect the two genders to have different characteristics. Perceptions of being treated poorly in the jobs in the various majors are found to be negatively correlated with the fraction of the people of one's own gender in the field of study, the wage gap, and beliefs of enjoying the coursework and working at the jobs. Finally, I find that academic performance is not the only consideration with regards to experimentation with majors.

On the methodology side, both chapters 2 and 4 add to the recent literature on subjective expectations (Manski, 2004). Chapter 2 contributes to this literature by providing an extensive description of students' expectations about major-specific outcomes, by using subjective expectations data to estimate a choice model, and by explaining the mechanisms through which beliefs form. Chapter 4 adds to the few studies in this literature that have looked at how individuals form and revise subjective expectations in response to new information. The panel on subjective beliefs allows me to answer several doubts that have been raised about the validity of subjective expectations data (Bertrand and Marianne, 2001). The results in chapter 4 bode well for the use of subjective expectations.

The analysis in chapter 3 is motivated by the finding in chapter 2 that individuals' college major choices are correlated with those of their parents and elder siblings. However, a positive correlation between an individual's choice of college major with that of his reference group is consistent with either the individual (1) learning about that particular choice through the experiences of others, and hence choosing that major (social learning), (2) getting a utility gain by simply having the same major as that of one's reference group (social comparison), or (3) sticking to the norm because of image-related concerns (social influence). Unfortunately, I cannot disentangle these mechanisms in my data. Moreover, though social interactions have been an active area of economic research for some time now, most studies focus on measuring the extent of social interactions and very little attention has been given to studying the mechanisms through which they are generated; this is primarily because of the various identification challenges that one faces when measuring social interactions (Manski, 1993, 2000). I tackle this issue in chapter 3 which outlines a simple model constructed on the premise that people are motivated by their own payoff and by how their action compares to others in their reference group. I show that conformity in actions may arise from learning about the norm (social learning or comparison concerns), or from image-related concerns (social influence). In order to empirically disentangle the two, I use the fact that image-related concerns can only be present if actions are publicly observable. The model predictions are tested in a charitable contribution experiment in which the actions and identities of the subjects are unmasked in a controlled and systematic way. The experimental setting provides an environment that provides clean evidence on each of these mechanisms, and also allows me to overcome the difficult identification problems in measuring social interactions in real world settings. I find that both learning about the norm and social influence play an important role in the choices of the subjects. Individuals indulge in social comparison and change their contributions in the direction of the social norm even when their identities are hidden. Once identities and contribution distributions of group members are revealed, individuals conform to the modal choice of the group. Moreover, social ties (defined as subjects knowing each other from outside the lab) affect the role of social influence. In particular, a low contribution norm evolves that causes individuals to contribute less in the presence of friends.

CHAPTER 2

College Major Choice and the Gender Gap

2.1. Introduction

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

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

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

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

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

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

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

Responses to questions eliciting subjective expectations match up with existing statistics for several questions indicating that respondents answer meaningfully and seriously. Respondents exhibit significant heterogeneity in their responses (both between and within

[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. The latter estimates a dynamic model of college and major choice under highly stylized assumptions on expectations formation.

genders), which underscores the importance of expectations data to conduct inference in settings with uncertainty. For example, the mean belief of being active in the full-time labor force at the age of 30 is 87.23% for females, and 95.11% for males. The gap widens for beliefs of labor force participation at the age of 40. Differences in beliefs could arise if people's experiences differ and beliefs are formed as a consequence of the individual's experiences and interactions with others in society. Other than that, beliefs could be shaped intentionally either by the subconscious, or by one's parents and peers. I find strong evidence of the latter- parents play a crucial role in shaping one's beliefs. Moreover, the effect differs by gender. For example, females with a stay-at-home mother have beliefs of being active in the full-time labor force at the age of 40 that are, on average, 12 points lower (on a 0-100 scale) than females with a working mother; no corresponding effect is found for males.

I estimate separate models for single major choice and for double major choice. The most important outcomes in the choice of single major are enjoying coursework, enjoying work at potential jobs, and approval of parents. Non-pecuniary outcomes explain about 45% of the choice behavior for males, and more than three-fourths of the choice for females. Males and females have similar preferences at college, but differ in their preferences regarding the workplace: males care more about the pecuniary outcomes in the workplace, females about the non-pecuniary outcomes. The results for the double major choice model are similar to those for single major. Graduating in 4 years, approval of parents, and enjoying coursework are the most important determinants of the choice. Additionally, I find evidence of individuals strategically choosing pairs of majors that allow them to specialize along certain dimensions. Females prefer pairs of majors which entail different chances of completion and getting a job upon graduation. On the other hand, males prefer major pairs that differ in their chances of completion, in the approval of parents, and in how much they would enjoy the coursework.

Besides being related to the literature on college major choice, this paper is related to three strands of literature. On the methodology side, it adds to the recent literature on subjective expectations (see Manski, 2004, for an overview of this literature). In the last decade or so, economists have increasingly undertaken the task of collecting and describing subjective data. Recently expectations data have been employed to estimate decision models. Van der Klaauw (2000) uses expectations data to improve the precision of the parameter estimates of a dynamic model of teacher career decisions. Delavande (2004) collects subjective data to estimate a choice model of birth control choice for women. The choice model used in this paper is motivated by her framework. The most recent step in this literature studies the formation of beliefs (Di Tella et al., 2007; and Lochner, 2007). My paper contributes to all three branches of this literature by providing an extensive description of students' expectations about major-specific outcomes, by using subjective expectations data to estimate a choice model, and by explaining the mechanisms through which beliefs form.

Second, this paper contributes to the recent literature on culture and economic outcomes (see Guiso et al., 2006; Alesina and Giuliano, 2007; Fernandez, 2007a). In order to establish a causal link from culture to economic outcomes, I focus on the dimension of culture that is inherited by an individual from previous generations, rather than being voluntarily selected. I use information on the country of origin of the individual's parents as a cultural proxy. Cultural proxies are found to bias beliefs in systematic ways, and the effect differs by gender. For example, after controlling for other factors, beliefs of females with foreign-born parents about being active in the labor force at age 30 are about 9 points lower than those of females with US-born parents; no such significant difference is found for males. I also find that cultural proxies bias preferences in favor of certain outcomes. Individuals with foreign-born parents value the pecuniary aspects of the choice more. In particular, males with foreign-born parents is the only sub-group in my sample for whom pecuniary outcomes explain more than 50% of the choice.

Finally, this paper is related to the literature that focuses on the underlying reasons for the gender gap in science and engineering. An interesting question is whether gender differences in choices are driven by differences in preferences or in beliefs. In the recent debate on the under-representation of women in science and engineering, some authors have claimed that the gap may be driven by the fact that women are less self-confident about their academic abilities than men. Valian (1998) argues that social prejudice against women causes them to lose self-confidence. Indeed, Solnick (1995) finds that women are more likely to shift to other majors from traditionally female majors if they attend a women's college. To check the validity of these hypotheses, I decompose the gender gap in major choice into differences in beliefs and differences in preferences. First, I find that gender differences in beliefs about ability constitute a small and insignificant part of the gap. This implies that explanations based entirely on the assumption that women have lower self-confidence relative to men (Long, 1986; Niederle et al., 2007) can be rejected in my data. Second, majority of the gender gap in majors that I consider can be explained by gender differences in beliefs about tastes for studying different fields, and preferences. For example, 60% of the gender gap in engineering is due to differences in preferences, while 30% is due to differences in how much females and males believe they will enjoy studying engineering. Gender differences in beliefs about future earnings in engineering are insignificant and explain less than 1% of the gap. I simulate an environment in which the female subjective belief distribution about ability and future earnings is replaced with that of males; in the case of engineering, this reduces the gap by about only 14%. These results suggest that simply raising expectations for women in science, as claimed by Valian (1998), may not be enough, and that wage discrimination and social biases may not be the main reason for why women are less likely to major in science and engineering.

The paper is organized as follows: Section 2.2 outlines the choice model and the identification strategy. Section 2.3.2 describes the institutional setup of Weinberg College of Arts & Sciences, outlines the data collection methodology, describes the subjective data, and discusses the formation of beliefs. Section 2.4 outlines the econometric framework used for estimation. Section 2.5 presents the estimation results for the single major choice model. Section 2.6 presents the results for the double major choice model. Section 2.7 undertakes a decomposition technique to understand the sources of gender differences in major choice. Finally, Section 2.8 concludes.

2.2. Choice Model

At time t, individual i is confronted with the decision to choose a college major from her choice set C_i. Individuals are forward-looking, and their choice depends not only on the current state of the world but also on what they expect will happen in the future. Individual i derives utility $U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it})$ from choosing major k. Utility is a function of a vector of outcomes $\mathbf{a}$ which are realized in college, a vector of outcomes $\mathbf{c}$ which are realized after graduating from college, and individual characteristics X_{it}. Examples of outcomes in **a** include graduating within 4 years, enjoying the coursework, and approval of parents. Examples of outcomes in **c** include future income, number of hours spent at the job, and ability to reconcile family and work. Both vectors, **a** and **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:

$$m \equiv \arg\max_{k \in C_i} \int U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it}) dP_{ikt}(\mathbf{a}, \mathbf{c}) \tag{2.1}$$

The goal is to infer the preference parameters from observed choices. However, the expectations of the individual about the choice-specific outcomes are also unknown. The most one can do is infer the decision rule conditional on the assumptions imposed on expectations. This would not be an issue if there were reason to think that prevailing expectations assumptions are correct. However, not only has the information processing rule varied considerably among studies of schooling behavior, most assume that individuals form their expectations in the same way.[3] First, there is little reason to think that

[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.

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

individuals form their expectations in the same way. Second, different combinations of preferences and expectations may lead to the same choice. Manski (2002) shows that different combinations of preferences and expectations (about others' behavior) leads to same actions 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 on which does not exist otherwise).

The exact utility specification is outlined in section 2.4 which presents the econometric framework. I first describe the data collection methodology in the following section.

2.3. Data

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

2.3.1. Institutional Details

At time t, the individual uses available information to form subjective beliefs $P_{ikt}(\mathbf{a}, \mathbf{c})$ $\forall k \in C_i$. She then uses her subjective beliefs and preferences to choose a major that maximizes her expected subjective utility. Over time she might acquire more information about any of the outcomes. For example, she may learn about her unobserved match quality (ability and taste) in different fields by taking courses. Moreover, she may also receive valuable information about the kinds of jobs and other major-related outcomes over time.

As shown in Figure A.3, the individual starts college at time 0 in her most preferred major. She may take courses in various majors between time 0 and time 1 in order to learn about her tastes and abilities. New information may arrive about match quality, or about the major-specific outcomes which could prompt the individual to change her major. She may switch her major any time between time 0 and time 1. At time 1, which corresponds to the end of the sophomore year, the individual has to declare her major. If she continues college after time 1, she takes further courses in her declared major, and graduates from college at time 2.

This goal is to estimate the individual's preferences between time 0 and time 1. Therefore, the study is restricted to Northwestern sophomores. Moreover, the model allows an individual to experiment with majors until time 1. I therefore restrict the study to schools at Northwestern where students have flexibility in choosing a major. For example, a student in the School of Engineering has to declare her major at time 0, and can only change her major by a special request to the school- she would not be eligible for the study. I further assume the choice set for an individual to be exogenous. This eliminates students in smaller schools at Northwestern since I will have to make strong assumptions about their choice set. Therefore, I restrict the study to the Weinberg College of Arts & Sciences (WCAS) at Northwestern. All sophomores with at least one major in the WCAS were eligible for the study.[4]

2.3.1.1. 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 order to limit the size of the choice set, I pool similar majors together. Table A.1 shows the

[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.

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. There is a trade-off between the number of categories and the length of the survey. This categorization is fairly fine, and also seems reasonable.

For a student pursuing a single major in WCAS, it is assumed that her choice set includes all the categories that span WCAS majors (a-g), and category k, the majors offered in the School of Engineering.[5] Therefore, any student with a single major is assumed to have 8 categories in her choice set.

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

2.3.2. 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

[5]This was done to elicit subjective beliefs of the outcomes associated with majoring in Engineering.

[6]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.

on campus in schools other than WCAS.[7] The E-mails and flyers explicitly asked for sophomores with an intended major in WCAS. Prospective participants were told that the survey was about the choice of college majors, and that they would get \$10 for completing the 45-minute electronic survey. It was emphasized that one need not have declared their major to participate in the study. The survey was conducted from November 2006 to February 2007. Respondents were required to come to the Kellogg Experimental Laboratory to take the electronic survey.

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

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

[7] 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).

Fine Arts. The proportion of females who (intend to) major in Literature and Fine Arts is more than 3 times the corresponding proportion of males.

The 45-minute survey consisted of three parts. The first part collected demographic and background information (including parents' and siblings' occupations and college majors, source of college funding etc.). The second part collected data relevant for the estimation of the choice model, and is discussed in more detail in the next subsection. The third part collected responses to open-ended questions intended to explore how respondents form expectations about various major-specific outcomes, and the sources of information they used. At the end of the survey, respondents were asked if they were willing to participate in a follow-up survey in a year's time.[8]

2.3.3. Subjective Data

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

a_1 successfully completing (graduating) a field of study in 4 years

a_2 graduating with a GPA of at least 3.5 in the field of study

a_3 enjoying the coursework

a_4 hours/week spent on the coursework

a_5 parents approve of the major

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

c_1 get an acceptable job immediately upon graduation

[8]If the respondent agrees to the follow-up, she is asked for her name and contact information. An astounding 97% (156 out of 161) respondents agreed to the follow-up.

c_2 enjoy working at the jobs available after graduation

c_3 able to reconcile work and family at the available jobs

c_4 hours/week spent working at the available jobs

c_5 social status of the available jobs

c_6 income at the available jobs

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

Note that $\{a_r\}_{r=\{1,2,3,5\}}$ and $\{c_q\}_{q=\{1,2,3\}}$ are binary, while outcomes a_4, and $\{c_q\}_{q=\{4,5,6\}}$ are continuous. For all $k \in C_i$, the following beliefs were elicited: $P_{ikt}(a_r = 1)$ for $r = \{1, 2, 3, 5\}$, $P_{ikt}(c_q = 1)$ for $q = \{1, 2, 3\}$, $E_{ikt}(a_4)$, and $E_{ikt}(c_q)$ for $q = \{4, 6\}$.

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

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

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

The questions dealing with subjective expectations were worded as follows:

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

and:

> *Look ahead to when you will be 30 YEARS OLD. If you majored in [X], what do you think is the percent chance that you will be able to*

reconcile work and your social life/ family at the kinds of jobs that will be available to you?

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

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

Social status of the available jobs was elicited as follows:

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

For the expected income, the question was as follows:[10]

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 full questionnaire can be viewed in Appendix A.1.

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

[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.

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

2.3.4. Data Description

Since the use of subjective data in economics is fairly recent, this section describes the subjective data in some detail. I discuss the precision and accuracy of the responses, and, whenever possible, compare them to objective measures. I also attempt to understand some of the determinants of beliefs; in particular, I study how beliefs for some outcomes are associated with family characteristics (as in Alesina and Giuliano, 2007). Readers interested in the model estimation may skip to section 2.4.

2.3.4.1. Subjective Beliefs of non-monetary outcomes. In order to highlight the heterogeneity in beliefs across respondents, I discuss the responses to two representative questions which elicit the subjective beliefs of choice-specific outcomes. Table A.4 presents the gender-specific subjective belief distribution of graduating with a GPA of at least 3.5 in Engineering, and Literature and Fine Arts, while Table A.5 shows the gender-specific distribution of the subjective probability of being able to reconcile work and family at jobs that would be available if one graduated in Social Sciences I, and Social Sciences II. Both tables show that respondents are willing to use the entire scale from zero to 100. It does seem that respondents tend to round off their responses to the nearest 5, especially for answers not at the extremes. There has been some concern that respondents might answer 50% when they want to respond to the interviewer but are unable to make any reasonable probability assessment of the relevant question.[11] However, the 50% response is not the most frequent one in the majority of the cases. There doesn't seem to be any

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

evidence of anchoring since numbers that were presented in the introductory text do not occur more often than others.

Table A.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.[12] 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. The different gender-specific belief distributions underscore the heterogeneity in beliefs between the two genders.

Analysis of Table A.5 also reveals substantial heterogeneity in responses. However, the gender-specific subjective distributions are similar in this case. Only a quarter of respondents believe the probability of being able to reconcile work and family at the jobs in Social Sciences II to be greater than 75%, while nearly 55% believe that to be the case at the jobs associated with graduating in Social Sciences I. These beliefs are consistent with the general perception of hectic work schedules in the corporate sector in which most Northwestern Social Sciences II undergraduates get jobs.

[12]Average GPA of Northwestern Engineering graduates of 2006 was 3.43, while that of Literature & Fine Arts was 3.56 (Source: Northwestern Graduate Survey). However, responses in Table A.4 also includes individuals who have *chosen* not to major in either of these two majors.

2.3.4.2. Subjective beliefs about Starting Salaries. Survey respondents were asked the average annual starting salary of Northwestern graduates of 2006 for various major categories. There were two reasons for asking this question. First, it allows me to check the plausibility of survey responses since they can be directly compared to actual salary realizations of 2006 graduates. Second, it allows me to gauge the respondents' level of knowledge about income differences across majors. The question asked was: *"What do you think was the average annual starting salary of Northwestern graduates (of 2006) with Bachelor's Degrees in Category X?"*. Though there's substantial heterogeneity in the empirical beliefs, I present average and median beliefs of respondents by gender in Table A.6. The first three columns show the actual outcomes for the 2006 graduating class. Females have lower average starting salaries across all major categories in WCAS (except Ethics and Values), and in most majors outside WCAS. The question posed to survey respondents asked for the average salary, so the point estimate that respondents provide could be a point on their subjective gender-specific earnings distribution, or the general earnings distribution. Since individuals majoring in a field may have better information about their chosen field, and may have beliefs different from those of individuals not majoring in it, I split survey responses by whether the respondent majors in the category about which the question is asked. Columns (4) and (5) present average and median beliefs of respondents who are pursuing a major in that category. In general, responses are consistent with actual trends. Relative magnitudes of responses for different majors match well with the actual statistics which shows that respondents are aware of different returns to majors. Males majoring in area studies overestimate the average earnings in the field. Female respondents overestimate average salaries for the three largest WCAS categories - Natural Sciences, Social Sciences I, and Social Sciences II.[13] The median and average responses for individuals not majoring in the field are shown in columns (6) and (7), and are remarkably close to the actual outcomes. On the whole, individuals seem to be well-informed about the differences in earnings across majors, and approximate the relative earnings reasonably well.

Using the demographic information collected from the respondents, one might be able to say something about the determinants of the errors in respondents' response to the question about salaries of 2006 graduates. To model the respondents' errors, I use the following metric:[14]

$$\ln \left| \frac{\widehat{s_{im}} - s_m^{obs}}{s_m^{obs}} \right|$$

where $\widehat{s_{im}}$ is respondent i's reported average starting salary in major m, and s_m^{obs} is the true average salary for Northwestern graduates of 2006 in major m. Column (1) of Table A.7 presents the results of regressing this metric for starting salaries in all majors on various demographic variables and a random effect to account for repeated observations for an individual. Column 2 (3) restricts the sample to cases where the respondents' point estimates are greater (less) than the observed outcomes. Individuals with higher GPAs make significantly larger errors when estimating starting salaries, and are more likely to overestimate them.[15] Females make larger errors than their male counterparts; moreover, females who overestimate (underestimate) make errors that are significantly

[13]This is the case when their responses are compared to either the average salaries for all graduates, or to those for females only.
[14]Betts (1996) uses this metric to examine undergraduates' errors in beliefs about salaries by type of education.
[15]This could be because such individuals think that GPA is a strong predictor of starting salary, when in fact GPA is not a significant predictor of one's starting salary in either the Northwestern Graduation Survey 2006, or the Baccalaureate & Beyond Longitudinal Study 1993/2003.

larger than those of males who overestimate (underestimate). In most specifications, individuals who have declared their major at the time of the survey, and whose parents attended college make smaller errors. The former observation is consistent with students who have declared their major being better-informed about the chosen field, while the latter is consistent with students with college-educated parents having access to better information. However, individuals with parents who have studied a given major are not better-informed about starting salaries in that major. Respondents who happen to be foreign students or second-generation immigrants are more likely to make larger errors.[16] Finally, respondents belonging to low-income households make smaller errors.[17]

Survey respondents were also asked the average salary they expect to earn at the age of 30 for each major category. There was substantial heterogeneity in responses. Table A.8 presents the average and median beliefs of the respondents. Unfortunately, Northwestern does not follow its alumni, and this data does not exist for previous graduate classes. For comparison purposes, I instead use the 2003 average annual salaries for 1993 college graduates from selective colleges in the Baccalaureate & Beyond Longitudinal Study (B&B: 1993/2003).[18] These statistics are presented in columns (1) and (2) of Table A.8. Again, the average and median beliefs of respondents majoring in the field are similar to those who do not major in that field. Both males and females report median and average salaries larger than those for the B&B sample (columns (1) and (2)). It could be that the

[16]Second-generation immigrants are defined as individuals who are US citizens, and have at least one parent who is foreign-born.

[17]This is in contrast to what Betts (1996) finds. This could be because the two studies survey individuals from different socio-economic backgrounds. Recall that the low-income category in my study is household income less than $150,000.

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

survey respondents are *self-enhancing* their own salary expectations.[19] However, there are at least three legitimate reasons why respondents' earning expectations may be different from the earnings statistics in the B&B sample. First, even though I have restricted the B&B sample to selective institutions, Northwestern graduates may work at jobs very different from those of graduates from comparable institutions. Second, respondents might think that future earnings distributions will differ from the current ones. Third, respondents may have private information (other than gender) about themselves which justifies having different expectations.

The discrepancy in the average and median responses for female respondents majoring in Natural Sciences, Social Sciences I, and Social Sciences II continues to be much larger than the corresponding discrepancy for other females and males. Given that the same females provided higher average responses for the starting salaries of 2006 graduates in these fields in Table A.7, it seems that they have misperceptions about actual outcomes.

2.3.4.3. Subjective Beliefs about Labor Force Participation. Beliefs of being active in the full-time labor force at the age of 30 and 40 were elicited from respondents. The median response for being active in the full-time labor force was same at both ages: 90% for females, and 95% for males. However, there is substantial heterogeneity in beliefs both between males and females, and within each gender group. Table A.9 shows the subjective belief distributions at the two ages.

The female subjective labor force distribution at the age of 30 is skewed to the left relative to the male distribution. Females have a lower mean belief about their labor force participation at the age of 30 than males (87.23% for females versus 95.11% for

[19]Smith and Powell (1990) find that male college seniors report higher income expectations for themselves than they do for their college peers at the same school.

males, with the gender difference significant at 0.01%). Moreover, females exhibit greater heterogeneity in their beliefs (a standard deviation of 13.56 for females versus 5.49 for males). Whereas nearly 80% of male respondents believe that there is a greater than 90 percent chance of their being active in the labor force at the age of 30, only 45% of females believe so.

The beliefs of being active in the full-time labor force at the age of 40 exhibit even greater heterogeneity between and within gender. The standard deviation of beliefs is 16.97 for females, and 7.57 for males. The mean belief for males is now 92.94%, and for females is 84.13%, with the gender difference being significant again. Now only about 65% of the males believe that the percent chance they will be active in the full-time labor force at the age of 40 is greater than 90%, while the corresponding number is 40% for females.

One can compare the median and mean beliefs of being active in the full-time labor force to a similar question in the expectations module of NLSY97. Though Northwestern undergraduates belong to a specific demographic, the comparison can still be useful. The question: *"What is the probability that you will be working for pay more than 20 hours per week when you turn 30?"* was posed to youth of ages 16-17 who are yet to start college (for details, see Fischhoff et al., 2000). The median response for both genders is 100%; the mean is 92.76% for males, and 91.84% for females. The difference in the mean belief between the NLSY97 females and those in my survey is significant (p-value = 0.016). Another statistic for comparison is the projected labor force participation for ages 25-34 in 2014. It is 95.3% for males, and 75.4% for females.[20] The mean for the male

[20]Source: U.S. Bureau of Labor Statistics, *Employment and Earnings*, January 2006.

respondents is very similar to the projected mean for the relevant age group, while the mean belief for females is about 10 percent points higher. Though females currently have a higher mean belief of being active in the labor force at 30 than the projected rate, their responses (relative to males) indicate that they start thinking about the uncertainty in their labor force status pretty early in their careers.

It might be of interest to see whether the heterogeneity across and within gender in beliefs about labor force participation can be explained by the demographic characteristics of the respondents. Table A.10 presents best linear predictors under square loss of the labor force participation rates. The belief of being active in the labor force for females is, on average, 6.7 (8.7) points lower than that of males at the age of 30 (40). Students with higher GPA have a higher belief of being active in the labor force at both 30 and 40. Individuals from higher income households have higher beliefs of being active in the labor force. Coefficients on parental education are not significant. McLanahan and Sandefur (1994) claim that children of divorced parents are more likely to be unemployed; however, in my sample, I don't find any such effect on the future labor force participation beliefs of individuals with divorced/separated parents. One notable finding is that individuals who are second-generation immigrants have a lower belief of being active in the labor force. A foreign-born parent decreases the belief of full-time labor force participation at the age of 40 by about 11.5 points for females, and 7 points for males. I treat country of birth of parents as a proxy for culture; since these individuals are born and raised in the US, they face the same institutions as individuals with US-born parents, but potentially differ in the cultural values transmitted to them by their parents. Focusing on the dimension of culture which is inherited by an individual (and hence exogenous) allows me to establish a causal link from culture to the economic outcome. Therefore, I conclude that culture is shaping individual's beliefs of labor force participation. This finding is similar to that of Fernandez and Fogli (2005) who find that cultural proxies have significant positive explanatory power for explaining work outcomes for second-generation American women (however, they use the female labor force participation rate in the female's country of ancestry as a cultural proxy).[21] Another significant finding is the effect of having a mother who is a full-time housewife on beliefs. Females with a stay-at-home mother have beliefs about labor force participation at the age of 40 which are, on average, 12.5 points lower than those of females with a working mother; no corresponding effect is found for males. In this context, it seems that beliefs for labor force participation for females are being shaped by the role of their mothers.[22]

2.3.4.4. Parents and Peer Effects. The importance of peer effects in shaping individual choices has been documented in several studies within higher education (see, for example, Betts and Morell, 1999), but there is little research on peer effects in crucial decisions such as choice of college major. Sacerdote (2001) does not find evidence for (roommate) peer effects in major choice for Dartmouth College roommates. De Girogi et al. (2007) find that Bocconi undergraduates are more likely to choose a major when many of their peers make that choice. Several respondents in my survey report to have majors that are the same as that of their roommates and friends. However, there is a

[21] Alesina and Giuliano (2007) also find that ancestry affects labor force participation of second generation immigrants.

[22] Fernandez (2007b) explains the s-shaped pattern observed in the female labor force participation in the last century in the US with an intergenerational learning model about payoffs to work for females; females receive private and public signals through which they learn about the payoffs to work. Here, it seems that females give a lot of weight to the signals they receive from their mothers. Also see Fogli and Veldkamp (2007) for a similar model where female labor force participation increases through learning from endogenous information.

self-selection issue: people often select with whom they associate.[23] Since rooming assignments are not totally random at Northwestern and there are endogeneity issues in how friendships are being formed, I cannot analytically study the strength of peer effects in the choice of college major.

Table A.11 presents the correlation patterns between the respondent's major and their father's major in Panel A, and the correlation pattern with the mother's majors in Panel B.[24] Since the sample is restricted to WCAS students, and several majors have been pooled together for each category, I cannot check for independence in the choice between an individual's choice and that of her parents. However, one feature that stands out is that students pursuing a major in Natural Sciences are more likely to have a parent who majored in that category. Moreover, of the 63 individuals with at least one sibling, 22 major in the same field as their sibling.

A positive correlation between an individual's choice of college major with that of her parents or siblings could be consistent with either (1) her having more information about that particular choice by information acquisition of the various outcomes from her parents and siblings, and hence choosing that major through an *indirect* effect of parents, (2) *direct* parental pressure leading an individual towards a particular major choice, or (3) a utility gain by studying the same major as that of parents. The first two are consistent with the evidence presented earlier. Moreover, when estimating preferences which incorporate individual heterogeneity in section 2.5.2, demographic characteristics (like country of birth of parents) are found to bias preferences for certain outcomes.

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

[24]Both majors of the individual are included in the table if they happen to pursue more than one major.

However, it is not possible to tell which mechanism is at work, i.e. whether beliefs and preferences are subconsciously being formed as a consequence of the individual's interactions with parents, or whether parents are intentionally shaping the beliefs and preferences of their children (as in Bisin and Verdier, 2001), or both. Survey respondents were asked to explain the reasons for the similarity between their major and that of their parents and siblings. Selected responses are shown in section A.2.2 of Appendix A. All three reasons come up as possible explanations. The responses also show instances of peer influence, but in most cases individuals seem to form friendships with similar individuals.

To conclude this section, I find that respondents provide meaningful answers to questions eliciting subjective expectations. In cases where responses could be compared to objective realities and statistics, survey responses match up well. Individuals are aware of the earnings differences across majors. However, females tend to make bigger errors about income expectations (overestimate future income), and seem to have misperceptions about future earnings in their own major in some cases. There is substantial heterogeneity in responses both between and within gender. This questions the accuracy of restrictions imposed on expectations in the literature. Since I don't observe the information set of the respondents, it is hard to pin down the exact mechanisms through which beliefs form. However, analysis of labor force participation beliefs and income expectations shows that beliefs for these specific outcomes are associated with culture and parents. Since I focus on aspects of culture (country of birth of parents; traits of parents) which are inherited by an individual, I can conclude that there is a causal link from culture and parents to beliefs about labor force participation.

2.4. Econometric Model

This section outlines the econometric framework.

Recall that utility, $U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it})$, is a function of a 5×1 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[25]; she chooses major m at time t if:

$$m \equiv \arg\max_{k \in C_i} \int U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it}) dP_{ikt}(\mathbf{a}, \mathbf{c}) \tag{2.2}$$

Moreover, as explained in section 2.3.3, 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. $\mathbf{d} = \{a_4, c_4, c_5, c_6\}$. The utility can now be written as a function of outcomes $\mathbf{b}$, $\mathbf{d}$, and characteristics X_{it}. I assume that utility is additively separable in the outcomes:

$$U_{it}(\mathbf{b}, \mathbf{d}, X_{it}) = \sum_{r=1}^{7} u_r(b_r, X_{it}) + \sum_{q=1}^{4} \gamma_{iqt} d_q + \varepsilon_{ikt} \tag{2.3}$$

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

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

up to the random term. Equation (2.2) can now be written as:

$$m \equiv \arg\max_{k \in C_i} \Big(\sum_{r=1}^{7} \int u_r(b_r, X_{it}) dP_{ikt}(b_r) + \sum_{q=1}^{4} \gamma_{iqt} \int d_q dP_{ikt}(d_q) + \varepsilon_{ikt} \Big) \tag{2.4}$$

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:

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

$$\Pr\begin{pmatrix} \sum_{r=1}^{7} \int u_r(b_r, X_{it}) dP_{imt}(b_r) + \sum_{q=1}^{4} \gamma_{iqt} \int d_q dP_{imt}(d_q) + \varepsilon_{imt} \\ \\ > \sum_{r=1}^{7} \int u_r(b_r, X_{it}) dP_{ikt}(b_r) + \sum_{q=1}^{4} \gamma_{iqt} \int d_q dP_{ikt}(d_q) + \varepsilon_{ikt} \end{pmatrix} \tag{2.5}$$

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

For the binary outcomes in $\mathbf{b}$, $P_{imt}(b_r)$ is simply $P_{imt}(b_r = 1)$ for $r \in \{1,..,7\}$; $P_{imt}(b_r = 1)$ 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 $\forall q \in \{1,..,4\}$.[26]

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

$$E_{ikt}(d_4) = \int w dG_{it}(w)[p_{ikt}E_{ikt}(I) + (1 - p_{ikt})E_{it}(I_0)] \qquad \text{for } k, p \in C_i \text{ and } p \neq k \tag{2.6}$$

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

where w is an indicator variable of the individual's labor force status, $G_{it}(w)$ is the subjective belief at time t about one's labor force status at the age of 30, and p_{ikt} is individual i's subjective probability at time t about successfully graduating in major k. The belief distribution of labor force status at the age of 30, $G_{it}(w)$, is simply $G_{it}(w=1)$; $\int wdG_{it}(w) = G_{it}(w=1)$, denoted as g_{it}, is elicited directly from the respondents.[27] 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.[28] Equation (2.5) can now be written as:

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

$$\Pr\left(\begin{array}{c}\sum_{r=1}^{7}\{P_{imt}(b_r=1)u_r(b_r=1,X_{it})+[1-P_{imt}(b_r=1)]u_r(b_r=0,X_{it})\}+\sum_{q=1}^{4}\gamma_{iqt}E_{imt}(d_q)+\varepsilon_{imt}\\ \\ >\sum_{r=1}^{7}\{P_{ikt}(b_r=1)u_r(b_r=1,X_{it})+[1-P_{ikt}(b_r=1)]u_r(b_r=0,X_{it})\}+\sum_{q=1}^{4}\gamma_{iqt}E_{ikt}(d_q)+\varepsilon_{ikt}\end{array}\right) \tag{2.7}$$

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

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

[28]In an earlier version of the model, I allow the individual to change fields of study once before dropping out of school. However, the results don't seem to change much.

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 $\triangle u_r(X_{it}) = u_r(b_r = 1, X_{it}) - u_r(b_r = 0, X_{it})$, 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:

$$\begin{aligned} & U_{imt}(\mathbf{b}, \mathbf{d}, X_{it}, \{P_{imt}(b_r = 1)\}_{r=1}^{7}, \{E_{imt}(d_q)\}_{q=1}^{4}) = \\ & = \sum_{r=1}^{7} P_{imt}(b_r = 1)\triangle u_r(X_{it}) + \sum_{r=1}^{7} u_r(b_r = 0, X_{it}) + \sum_{q=1}^{4} \gamma_{iqt} E_{imt}(d_q) + \varepsilon_{imt} \end{aligned} \tag{2.8}$$

Equation (2.7) can now be written as:

$$\begin{aligned} & \Pr(m | X_{it}, \{P_{ikt}(b_r), E_{ikt}(d_q)\}_{r\in\{1,..,7\}, q\in\{1,..,4\};\, k\in C_i}) = \\ & \Pr\begin{pmatrix} \sum_{r=1}^{7} P_{imt}(b_r = 1)\triangle u_r(X_{it}) + \sum_{q=1}^{4} \gamma_{iqt} E_{imt}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^{7} P_{ikt}(b_r = 1)\triangle u_r(X_{it}) + \sum_{q=1}^{4} \gamma_{iqt} E_{ikt}(d_q) + \varepsilon_{ikt} \end{pmatrix} \\ & \forall k \in C_i,\ m \neq k \end{aligned} \tag{2.9}$$

$\{\triangle u_r(X_{it})\}_{r=1}^{7}$, and $\{\gamma_{iqt}\}_{q=1}^{4}$ are the parameters to be estimated. g_{it}, $\{P_{ikt}(b_r = 1)\}_{r=1}^{7}$, and $\{E_{ikt}(d_q)\}_{q=1}^{3}$, 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 random terms required for identifying the model parameters are discussed in the next section.

2.5. Single Major Choice Model

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

2.5.1. Estimation with Homogenous Preferences

The model described in section 2.4 assumes that the utility function for the binary outcomes $u_r(b_r, X_i)$, and the constants on continuous outcomes ($\{\gamma_{iq}\}_{q=1}^{4}$) depend on individual characteristics. I initially assume that the utility function does not depend on individual characteristics. Under this assumption, (2.9) becomes:

$$\Pr(m|P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\ k\in C_i})$$

$$= \Pr\begin{pmatrix} \sum_{r=1}^{7} P_{im}(b_r=1)\triangle u_c + \sum_{q=1}^{4}\gamma_q E_{im}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^{7} P_{ik}(b_r=1)\triangle u_c + \sum_{q=1}^{4}\gamma_q E_{ik}(d_q) + \varepsilon_{ikt} \end{pmatrix} \tag{2.10}$$

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

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

$$\Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\ k\in C_i}) \tag{2.11}$$

$$= \frac{\exp(\sum_{r=1}^{7} P_{im}(b_r=1)\triangle u_r + \sum_{q=1}^{4}\gamma_q E_{im}(d_q))}{\sum_{k\in C_i}\exp(\sum_{r=1}^{7} P_{ik}(b_r=1)\triangle u_r + \sum_{q=1}^{4}\gamma_q E_{ik}(d_q))}$$

Under these parametric assumptions, the parameters $\{\triangle u_r\}_{r=1}^{7}$, and $\{\gamma_q\}_{q=1}^{4}$are identified. The elicited subjective probabilities described in section 2.3.2 are used in estimation.

Column (1) of Table B.1 presents the maximum likelihood estimates using stated choice data.[29,30]

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

[29] 44 of the 83 respondents with a single major had declared their major at the time of the survey. For the remaining 39, I use their stated intended choice for estimation.

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

Column (2) of Table B.1 shows the maximum-likelihood estimates based on (2.11) with the addition of female interactions in order to get some measure of relative differences between males and females. For males, the difference in utility levels is largest for enjoying coursework, finding a job upon graduation, and the social status of the jobs in decreasing order of importance. For females, the three outcomes that matter the most are graduating in 4 years, enjoying the coursework, and enjoying work at the available jobs. Though income stays insignificant, the coefficient on income interacted with the female dummy shows that the negative coefficient on income in Column (1) is being driven by the preferences of females; income has a positive coefficient for males now, and negative for females (though neither are significant).

In addition to stating their choice, respondents were also asked to rank the elements in their choice set. The stated preference data provides more information which can be used for estimation of the model parameters. Under the assumptions of standard logit, the probability of any ranking of alternatives can be written as a product of logits. For example, consider the case where an individual's choice set is $\{a, b, c, d\}$. Suppose she ranks the alternatives b, d, c, a from best to worst. Under the assumption that the 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:[31]

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

(2.12)

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

Column (3) in Table B.1 presents the maximum likelihood estimates using stated preference data. The difference in utility levels is still largest and positive for enjoying coursework. Graduating in 4 years, the second most important outcome using stated choice data, is now negative but not significant. Enjoying work at the jobs is the second most important outcome with a positive coefficient. Approval of parents, now the third most important outcome, has a positive coefficient that is one-half that of enjoying coursework. The difference in utility levels for graduating with a GPA of at least 3.5 is now positive and significant. Status of the jobs continues to be important: a unit increase in the social status of the jobs changes the utility by as much as a 4% increase in the probability of enjoying coursework. The difference in utility levels for other binary outcomes is not significantly different from zero. The coefficient on income is now positive, but not significant.

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

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

In order to get a measure of the magnitude of the e stimated parameters, the natural thing would be to do willingness to pay calculations, i.e. translate the differences in utility levels into the amount that an individual would be willing to forgo at the age of 30 in earnings in order to experience that outcome.[32] However, since expected income at age 30 is not significant in any of the specifications considered, the standard errors on such calculations are huge, and the results are not very meaningful. I, therefore, don't present the willingness to pay calculations. Instead, 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,

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

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:

$$(2.13) \qquad M_{X_1}$$

$$\equiv \quad \| \overline{\Pr(choice \ = j| \ \{\widehat{\beta_1}, \widehat{\beta_2}\}} - \overline{\Pr(choice \ = j| \ \{\widehat{\beta_1} = 0, \widehat{\beta_2}\}} \|$$

$$= \quad \sqrt{\sum_{j=1}^{8} \left[\sum_{i=1}^{N} \frac{\Pr(choice \ = j| \ \{\widehat{\beta_1}, \widehat{\beta_2}\})}{N} - \sum_{i=1}^{N} \frac{\Pr(choice \ = j| \ \{\widehat{\beta_1} = 0, \widehat{\beta_2}\})}{N} \right]^2}$$

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

$$(2.14) \qquad M_{X_2}$$

$$\equiv \quad \sqrt{\sum_{j=1}^{8} \left[\sum_{i=1}^{N} \frac{\Pr(choice \ = j| \ \{\widehat{\beta_1}, \widehat{\beta_2}\})}{N} - \sum_{i=1}^{N} \frac{\Pr(choice \ = j| \ \{\widehat{\beta_1}, \widehat{\beta_2} = 0\})}{N} \right]^2}$$

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 B.2 presents the results of this decomposition strategy. Each cell shows the *relative* contribution (R) of the outcome to the choice. Panel B of Table B.2 presents the results of this decomposition technique using the estimates obtained from stated preference data. Column (1) shows the decomposition results of the estimates of the pooled sample: nearly three-fourths of the choice is driven by the non-pecuniary outcomes.[33] If the decomposition is made finer, one can see that parent's approval and enjoying coursework jointly explain about 45% of the choice. Pecuniary outcomes associated with college (hours/week spent on coursework, graduating with a GPA of at least 3.5, and graduating in 4 years), and workplace (finding a job upon graduation, hours/week spent at work, income at the age of 30, and the social status of the jobs) each account for about 20% of the choice.

The estimates of the pooled sample mask the differences between males and females. Columns (2) and (3) of Table B.2 show the decomposition results using the estimates from the male sub-sample, and the female sub-sample respectively. Non-pecuniary outcomes explain about 45% of the choices for males, but more than 80% of the choice for females. Parent's approval and enjoying coursework are the most important outcomes for females explaining about 45% of their choice, while pecuniary outcomes associated with the workplace are of utmost importance to males explaining 48% of their choice. Reconciling family and enjoying work at the available jobs are second in terms of importance to females, but of least importance to males. On the whole, non-pecuniary determinants are crucial in explaining the choices for both males and females. However, males and

[33]Outcomes classified as being non-pecuniary are: parent's approval, enjoying coursework, reconciling work and family, and enjoying work at the jobs. The remaining outcomes are termed as being pecuniary.

females differ in their preferences in the workplace: males value pecuniary aspects of the workplace more, while females value non-pecuniary aspects of the workplace more.

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

Before I conclude the discussion of the homogenous choice model, I discuss some robustness checks that I did in order to figure out whether income is actually insignificant in the choice of major, or if the result is driven by large standard errors. The descriptive analysis of respondents' expectations of income in different majors in Table A.8 indicates that students are aware of the income differences across majors, but 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 (2.14) for 1000 bootstrap samples for each of the sub-samples. The bootstrap confidence interval of R_{γ_4} for both males and females does not include zero: the higher end of the 90% bootstrap interval for expected income is 16% and 7.5% for males and females respectively. This seems to suggest that γ_4 is insignificant because of a large standard error, and not because it is a precise zero.[34]

2.5.2. Estimation with Heterogeneous Preferences

The analysis undertaken in section 2.3.4 shows that beliefs for various outcomes are associated with demographic characteristics and cultural proxies. However, it could be the case that preferences for the different outcomes also depend on individual characteristics. For example, if individuals have declining marginal utility of consumption, and preferences are separable in consumption and non-pecuniary outcomes, then the value of pecuniary outcomes will be higher for individuals from low-income households. Such heterogeneity, if not accounted for, may bias the estimates presented in section 2.5.1. Several empirical studies have documented the influence of family and society in the endogenous formation of preferences. For example, Fernandez et al. (2004) find that whether a male's mother worked while he is growing up is correlated with whether his wife works, and interpret this as preference transmission. Moreover, Guiso et al. (2006) present evidence of culture affecting individuals' preferences.[35] I now relax the assumption of section 2.5.1

[34] An additional robustness check that I did was to estimate the model using the ordinal ranking of income (instead of expected income). This allows me to control for the noise in the reported income expectations. The coefficient on (ranked) income is now significant for the males, but continues to be insignificant for females. Moreover, the confidence interval of R_{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 towards the choice (ranked income and status are highly correlated). Therefore, none of the results change. However, this seems to suggest that income is at least significant for males.

[35] Also see Doepke and Zilibotti (2007); their theoretical framework of occupational choice models culture as a feature of preferences.

that the utility for each binary outcome $u_r(b_r)$, and the constants γ_q for the continuous outcomes do not depend on individual characteristics other than gender. Though I have relatively rich demographic information on the respondents, it is not possible to account for heterogeneity in all outcomes because of the small sample size. I, therefore, consider heterogeneity along the following dimensions:

(1) An individual might care about her parent's approval for several reasons. She might be more inclined to ensure that her parents approve of her choice if she relies on them for college support. Moreover, concern for parent's approval might depend on the individual's cultural and ethnic background. I allow for heterogeneity in the utility for approval of parents by incorporating the financial support an individual receives from her parents when in college, and whether her parents are foreign-born or not.

(2) Children growing up in divorced/separated households make different choices than other individuals (Gruber, 2004). Here I consider the effect of growing up in such a household on the individual's preference for being able to reconcile work and family.

(3) An individual's preference for the social status of jobs may vary by her cultural background. In certain cultures, immense importance is given to the status of the jobs. This heterogeneity is accounted for by taking into account whether the individual's parents are foreign-born.

(4) If non-pecuniary outcomes are a normal good, an individual from a low-income family will value the income profiles associated with the majors more relative to other individuals. I account for this heterogeneity by including information on

parent's annual income. I also allow for heterogeneity by taking into account whether an individual's parents are foreign-born or not.

The enriched utility function for individual i is:

$$U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r\in\{1,.,7\},q\in\{1,.,4\}})$$

$$= \sum_{r=\{1,2,3,5,6,7\}} P_{im}(b_r = 1)\triangle u_r + \triangle u_4[\textit{parents'\_support}_i \times (1\text{-}\textit{Foreign}_i) \times P_{im}(b_4 = 1)]$$

$$+\widetilde{\triangle u_4}\ [\textit{parents'\_support}_i \times \textit{Foreign}_i \times P_{im}(b_4 = 1)] + \widetilde{\triangle u_7}\ \ [\textit{divorced}_i\ \times\ P_{ijt}(b_7 = 1)]$$

$$+\textstyle\sum_{q=1}^{2} \gamma_q E_{im}(d_q) +\ \gamma_3\ [(1\text{-}\textit{Foreign}_i)\ \times E_{im}(d_3)]\ +\ \widetilde{\gamma_3}\ [\textit{Foreign}_i\ \times E_{im}(d_3)] + \gamma_4 E_{im}(d_4) \tag{2.15}$$

$$+\ \gamma_4^{HI}[E_{im}(d_4) \times\ (1\text{-}\textit{low\_inc}_i) \times (1\text{-}\textit{Foreign}_i)]\ \ +\ \widetilde{\gamma_4^{HI}}[E_{im}(d_4) \times\ (1\text{-}\textit{low\_inc}_i) \times \textit{Foreign}_i]$$

$$+\ \gamma_4^{LI}[E_{im}(d_4) \times\ \textit{low\_inc}_i \times (1\text{-}\textit{Foreign}_i)]\ +\ \widetilde{\gamma_4^{LI}}[E_{im}(d_4) \times \textit{low\_inc}_i \times \textit{Foreign}_i]\ + \varepsilon_{im}$$

$$\forall\ m = 1, ., 8$$

where *low_inc* is a dummy variable that equals one if the individual's parents earn less than \$150,000 annually; *parents'_support* captures the financial support an individual receives from her parents,[36] *Foreign* is a dummy that equals one if either of the individual's

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

parents is foreign-born, and *divorced* is a dummy that equals one if the individual's parents are either separated or divorced.

I continue to assume that the random terms $\{\varepsilon_{ik}\}$ are independent for every individual i and choice k. Table B.4 presents the maximum likelihood estimates of this model using stated preference data. Estimates from the pooled sample in Column (1) show that difference in utility levels is still largest and positive for enjoying coursework, and that the coefficient is almost unchanged from the specification with homogenous preferences. The coefficients of the outcomes for which heterogeneity is not considered stay almost the same as that in the earlier specification. With this enriched specification, the difference in utility levels for parent's approval is 0.34 for individuals with US-born parents who do not receive college support from their parents, and 2.04 for individuals who annually receive more than \$25,000 in college support from their parents. This is consistent with the hypothesis that approval of parents matters more to individuals who depend on their parents for college funding. However, I don't find support for this hypothesis for individuals with foreign-born parents. The difference in utility levels for reconciling work and family continues to be insignificant. Individuals with separated or divorced parents have a negative coefficient for reconciling work and family, but it is not significantly different from zero. Introducing heterogeneity for the status outcome gives an interesting result. Status of the available jobs, an important determinant in the choice in the earlier specifications, is not important to individuals with US-born parents. However, for individuals with foreign-born parents, a unit increase in the social status of the jobs changes the utility by as much as a 8% increase in the probability of the most important outcome, enjoying coursework. This implies that the large positive coefficient on the social status of jobs in earlier specifications is being driven by the preferences of individuals with foreign-born parents in the sample. The coefficient on income at age 30 is still not significantly different from zero. However, there is weak support for the hypothesis that individuals from low-income households value the future earnings profile more in their choice.

Columns (2) and (3) of Table B.4 present the results of the heterogeneous choice model for the male and female sub-sample respectively. Coefficients of outcomes which are not interacted with any demographic variables are almost unchanged with respect to the corresponding specification (column 4 in Table B.1). For males with US-born parents, difference in utility levels for approval of parents varies from 0.578 when receiving no support from parents to 3.47 when annually receiving more than $25,000 in support from them. The corresponding coefficient for females with US-born parents is only half in magnitude to that for males. The coefficient on parents' approval for females with foreign-born parents is similar in magnitude to that of males with US-born parents. Surprisingly, the utility change in approval of parents for males with foreign-born parents is not significantly different from zero. On the other hand, social status of jobs only matters to males with foreign-born parents: a unit increase in the social status of the jobs changes the utility by about a 13% increase in the probability of enjoying coursework for these males. Earnings at the age of 30 are a significant determinant for males belonging to low-income families with foreign-born parents.

To gain insight into the magnitude of these parameters, Table B.5 shows the results of the decomposition methodology outlined in equation (2.14). Except for males with foreign-born parents, non-pecuniary attributes explain more than half of the choice. For individuals with US-born parents, more than two-thirds of the choice is driven by non-pecuniary motivations; the non-pecuniary outcomes at college are of utmost importance to this group. For individuals with foreign-born parents, pecuniary outcomes at the workplace are of greatest value in the choice for males, while non-pecuniary outcomes at college continue to be of utmost importance to such females.

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

2.5.3. Parent's Approval

The estimation results in sections 2.5.1 and 2.5.2 show that approval of parents is an important determinant in the choice for males with US-born parents and for all females. The social psychology literature documents a similar finding for females: Vincent et al. (1998) find that females' perceptions of their parent's preferences for them predict their career orientation. Though section 2.5.2 shows that one channel through which parent's approval matters is the individual's reliance on them for college support, it is not clear what majors parents are more likely to approve, and what criteria they use for approving a major. Since only the beliefs of students are observed, I can only study the relationship between students' beliefs of parent's approval of a major and their own beliefs of other outcomes associated with the choice.[37] Controlling for the individual's major, I regress respondent i's beliefs about her parent's approval for major j on her beliefs about the other outcomes associated with j. More specifically, I consider the following regression model:

$$P_{ij}(b_4 = 1) = \delta_i + \lambda_j + \boldsymbol{\alpha}' X_{ij} + \boldsymbol{\beta}' \left[\sum_{\substack{c=1 \\ c \neq 4}}^{7} P_{ij}(b_c = 1) + \sum_{q=1}^{4} E_{ij}(d_q) \right] + \varepsilon_{ij} \tag{2.16}$$

where δ_i is an individual fixed-effect, λ_j is a field-fixed effect, X_{ij} is a vector of individual-specific controls, and $\boldsymbol{\beta}$ is the vector of interest. The results are presented in Table B.6. Students' beliefs of parent's approval for a given major increase in their beliefs of finding a job upon graduation, enjoying work at potential jobs, and social status of jobs. Expectation of parent's approval for a major increases by nearly 3 points (on a scale of 0 -100) if the probability of finding a job upon graduation in that major increases by 10 points. This effect is even stronger for students with foreign-born parents: students believe that switching to a major with a 10 points higher probability of getting a job upon graduation is likely to increase parent's approval by nearly 5 points. A positive and significant effect, half in magnitude to that of finding a job, is found for the social status of the jobs. Again the effect is stronger for students with foreign-born parents. The only other outcome that affects beliefs about parent's approval is the expectation of enjoying work at the jobs for females.

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

Males with foreign-born parents expect approval of parents for a major to increase by about 12.5 points for a unit increase in the social status of the jobs. This result reconciles the earlier finding in section 2.5.2 of parent's approval not mattering to males with foreign-born parents. Expectation of parent's approval has a positive relationship with the perceived social status of jobs, and status of jobs is an important outcome only in the choice for males with foreign-born parents (column (2) in Table B.4); hence, because of colinearity, approval of parents does not *directly* affect the choice of these individuals.

2.5.4. Robustness Checks

The model estimated in section 2.5.1 assumes that all individuals have homogeneous preferences for various outcomes. Individuals with different characteristics are very likely to have different preferences. Moreover, the assumption that the random terms $\{\varepsilon_{ik}\}$ are independent for every individual i and choice k might be very strong. Though a model with limited heterogeneity in preferences is estimated in section 2.5.2, any unaccounted or unobserved heterogeneity may bias the model estimates. In this section, I specify a random parameters logit model to account for these issues (see Revelt and Train, 1997, for a discussion of mixed logit models). One could allow heterogeneity in preferences for all outcomes, but I focus on the most important outcomes: I consider a model in which the differences in utility levels for graduating with a GPA of at least 3.5, enjoying the coursework, approval of parents, enjoying work at the available jobs, and the parameter for social status of the available jobs are allowed to vary in the population with a specified distribution. The utility that individual i receives from choosing major m is:

$$
\begin{aligned}
& U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\}}) \\
& = \textstyle\sum_{r=\{1,5,7\}} P_{im}(b_r = 1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s = 1)\triangle u_{si} \\
& + \textstyle\sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3) + \varepsilon_{im}
\end{aligned}
\tag{2.17}
$$

where $\triangle u_{si}$ for $s = \{2, 3, 4, 6\}$, and γ_{3i} are allowed to vary in the population according to a specified parametric distribution, and ε_{im} is an iid random term that is extreme value distributed. I denote the vector of parameters $\{\triangle u_{2i}, \triangle u_{3i}, \triangle u_{4i}, \triangle u_{6i}, \gamma_{3i}\}$ by $\boldsymbol{\beta}_i$, and the density of these parameters $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the parameters of the distribution. The probability of i choosing the major m *conditional* on $\boldsymbol{\beta}_i$ is:

$$
\begin{aligned}
& \Pr(m|\boldsymbol{\beta}_i) = \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\ k\in C_i}, \boldsymbol{\beta}_i) = \\
& = \frac{\exp(\sum_{r=\{1,5,7\}} P_{im}(b_r = 1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s = 1)\triangle u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3))}{\sum_{k\in C_i} \exp(\sum_{r=\{1,5,7\}} P_{ik}(b_r = 1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{ik}(b_s = 1)\triangle u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{ik}(d_q) + \gamma_{3i} E_{ik}(d_3))}
\end{aligned}
\tag{2.18}
$$

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

$$
P_{im}(\theta) = \int \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\ k\in C_i}, \boldsymbol{\beta}_i)\, f(\boldsymbol{\beta}_i|\boldsymbol{\theta}) d\boldsymbol{\beta}_i
\tag{2.19}
$$

This integral is approximated through simulation since it cannot be calculated analytically. For a given value of the parameter vector $\boldsymbol{\theta}$, a value of $\boldsymbol{\beta}_i$ is drawn from its distribution.

Using this draw, the conditional probability is calculated. This process is repeated for D draws, and the average is taken as the approximate choice probability:

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

The log likelihood function $\sum_i \ln(\Pr_i)$ is approximated by the simulated log-likelihood function $\sum_i \ln(\widehat{P_i(\theta)})$, and the estimated parameters are those that maximize the simulated log-likelihood function. I assume that the coefficients for graduating with a GPA of at least 3.5, enjoying the coursework, approval of parents, enjoying work at the available jobs, and social status of the available jobs are independently log-normally distributed.[38] The difference in utility levels for an outcome k which is assumed to vary in the population is expressed as $\triangle u_k = \exp(\overline{\triangle u_k} + \sigma_k \mu_k)$ where μ_k is a standard normal deviate. The parameters $\overline{\triangle u_k}$ and σ_k, which represent the mean and standard deviation of $\log(\triangle u_k)$ are estimated. The mean and standard deviation of $\triangle u_k$ are $\exp(\overline{\triangle u_k} + \frac{\sigma_k^2}{2})$ and $\exp(\overline{\triangle u_k} + \frac{\sigma_k^2}{2}) * \sqrt{(\exp(\sigma_k^2) - 1)}$ respectively.

Columns (1a)-(1c) in Table B.7 present the estimates of the mixed logit specification for the model with $D = 100,000$. Estimates of various outcomes are similar to those obtained in the corresponding model with no heterogeneity (column (3) of Table B.1). The mean coefficient of enjoying coursework is still largest in absolute value and significant. The estimated standard deviations of the (random) coefficients are highly significant indicating that these parameters do indeed vary in the sample. Standard deviations for

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

coefficients of graduating in 4 years and social status of available jobs are especially very large, indicating that there is substantial heterogeneity in how these outcomes are valued in the sample (consistent with what was also found in the previous section). Another point of note is that the mean coefficients in the mixed logit model are larger than the corresponding fixed coefficients in Table B.1. This is because in the mixed logit, some of the stochastic portion of the utility is captured in $\boldsymbol{\beta}_i$ rather than in ε_i. Since the utility is scaled so that ε_i has the variance of an extreme value, the parameters are scaled down in the standard model relative to the mixed logit model (the same result is obtained by Revelt and Train, 1998, and Brownstone and Train, 1999). The fact that the mean coefficients are bigger than the fixed coefficients implies that the random parameters constitute a large share of the variance in unobserved utility.

One might wonder as to what extent can the variation in the parameters in the mixed logit model be explained by including demographic characteristics. Columns (2a) through (2c) in Table B.7 present estimates of the mixed logit model with demographic variables that were used in the heterogeneous model described in section 2.5.2. The estimates are similar to those in column (1) of Table B.4, though they are larger in magnitude, which is expected. The standard deviations are still large and significant which indicates that the demographic variables considered in section 2.5.2 only capture some of the heterogeneity that is exhibited by the individuals. Nonetheless, the fact that the relative magnitude of the estimates is similar to previous results is reassuring.

2.6. Double Major Choice Model

For reasons that will become clear shortly, a separate choice model is estimated for double majors. Nearly half of the sample respondents state that they are pursuing two majors. Anecdotal evidence suggests that about half of them will end up dropping one of their majors some time before graduation.[39] Since I have stated preference data from these respondents, I first estimate the same model as in section 2.5.1 in order to get a sense of the motivations of the choice for these individuals. The parameter estimates (available upon request) are similar to those for respondents pursuing a single major. Table B.8 presents the decomposition results of equation (2.14) using these estimates. As before, non-pecuniary attributes explain most of the choice. It does seem that these individuals are similar to those pursuing single majors in their preferences for various outcomes.

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

2.6.1. Estimation of Double Major Choice Model

Depending on the exact composition of the individual's major pair, the choice set of the individual now consists of either 8 or 9 categories.[40] For estimation, I assume that the individual may choose a single major or a pair of majors. The set of alternatives available to the individual includes all subsets of two majors in WCAS (${}^{8}C_{2} = 28$), all possible single majors in WCAS (7), and all possible pairings of WCAS majors with non-WCAS

[39]According to the Registrar's Office and Northwestern Graduation Survey 2006, less than 30% of WCAS undergraduates graduate with more than one major.

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

majors for a total of 70 alternatives. The distribution of majors for individuals pursuing double majors in the sample is shown in Table B.9. There's no obvious pattern in which individuals are choosing pairs.

The major-specific outcomes that appear in the utility function remain the same as before, but the form of the utility function is now different. Before specifying the structural form of the utility function, it may be useful to think about why an individual may decide to choose two majors. Respondents pursuing more than one major were asked to explain reasons for pursuing more than one major; selected responses are shown in Appendix A (section A.2.1). Two main reasons emerge: first, two majors appropriately differentiated can provide a broader mix of options than a single major; second, it might be the case that no single major meets the needs of the individual. For example, an individual might be interested in both maximizing her income prospects as well as enjoying the coursework. It could very well be the case that no single major meets her needs, but a combination of two majors does. To capture the enhanced options and specialization of function that two majors provide, I assume that the utility of a major pair depends on the attributes of each major separately, and on the attributes of a composite major combining the best of both majors. However, I only apply the idea of a composite major to outcomes associated with college. Outcomes associated with the workplace are not considered since they come as a package; for example, one does not have the option to choose the income associated with the jobs available in one major, and the lifestyle associated with the jobs in the second major. I also do not consider the composite major representation for graduating with a GPA of more than 3.5 because GPA is a composite of all coursework an individual does, and it is not possible to hedge along this dimension. The outcomes for which the composite major specification is used are: graduating in 4 years, hours/week spent on coursework, enjoying the coursework, approval of parents, and finding a job upon graduation. The utility function of a major pair p consisting of majors p_1 and p_2 takes the form:

$$\begin{aligned} U_{ipt} &= U_{ip_1t}(\mathbf{b},\mathbf{d},X_{it},\{P_{ip_1t}(b_r=1)\}_{r=1}^{7},\{E_{ip_1t}(d_q)\}_{q=1}^{4}) \\ &+U_{ip_2t}(\mathbf{b},\mathbf{d},X_{it},\{P_{ip_2t}(b_r=1)\}_{r=1}^{7},\{E_{ip_2t}(d_q)\}_{q=1}^{4}) \qquad (2.20) \\ &+U_{i\widetilde{p}t}(\mathbf{b},\mathbf{d},X_{it},\sum_{r=\{1,3,4,5\}}\max[P_{ip_1t}(b_r=1),P_{ip_2t}(b_r=1)],\ \max[E_{ip_1t}(d_1),E_{ip_2t}(d_1)]) \end{aligned}$$

where $U_{ip_1t}(.)$ is as defined in equation (2.8), and $\widetilde{p}$ refers to the composite major. Since there is no way of specifying a "primary" and a "secondary" major, I use the same functional form for the utility of each major in one's major pair, i.e. $U_{ip_1t} = U_{ip_2t}$. Since $U_{ip_1t}(.)$ is linear-in-parameters, the average characteristics of the two majors appear in the utility function. Assuming that the utility function does not depend on the individual characteristics, X_{it}, and dropping the time subscript, the utility function can be written as:

$$
\begin{aligned}
&U_{ip}(\{P_{ip_1}(b_r), E_{ip_1}(d_q), P_{ip_2}(b_r), E_{ip_2}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\}}) \\
&= \sum_{r=1}^{7}\{\tfrac{P_{ip_1}(b_r=1)+P_{ip_2}(b_r=1)}{2}\}\triangle u_{r1} + \sum_{q=1}^{4}\gamma_{q1}\{\tfrac{E_{ip_1}(d_q)+E_{ip_2}(d_q)}{2}\} \\
&+ \sum_{r=\{1,3,4,5\}} \max[P_{ip_1}(b_r=1), P_{ip_2}(b_r=1)]\triangle u_{r2} \\
&+\gamma_{12}\min[E_{ip_1}(d_1), E_{ip_2}(d_1)] + \varepsilon_{ip} \;=\; U_{ip} \;+\; \varepsilon_{ip}
\end{aligned}
\tag{2.21}
$$

The composite major representation captures the notion of functional specialization as follows: say an individual with a major pair chooses one major with a low completion probability because of some of its other attributes, and a second major where the completion probability is the most important consideration. Given the specification above, one would expect $\triangle u_{11} \approx 0$ and $\triangle u_{12} > 0$ in this case of extreme specialization. On the other hand, for an individual who equally values the completion probabilities associated with both her majors, one would expect $\triangle u_{12} \approx 0$ and $\triangle u_{11} > 0$. Thus the ratio $\triangle u_{12}/\triangle u_{11}$ ($\{\triangle u_{r2}/\triangle u_{r1}\}_{r=\{1,3,4,5\}}$, γ_{12}/γ_{11}) is a measure of the extent to which an individual desires to functionally specialize her majors along the given outcome.

I continue to assume that the random terms $\{\varepsilon_{ip}\}$ are independent for every i and every p, and have a extreme value distribution. The maximum likelihood estimates are shown in Table B.10. Panel B shows $\{\triangle u_{r1} + \triangle u_{r2}\}_{r=1}^{7}$; the relative magnitudes of $\{\triangle u_{r1} + \triangle u_{r2}\}_{r=1}^{7}$ are a measure of the importance of each outcome in choosing a major pair. For the pooled sample results presented in column (1), the difference in utility levels for graduating in 4 years is positive and largest in magnitude. The next most important outcome is enjoying the coursework with a positive coefficient. Approval of parents is the third most important outcome. Enjoying work at potential jobs, graduating with a GPA of at least 3.5, finding a job, and reconciling work and family are next in order of importance. All four are significant with positive coefficients. The coefficient on hours/week at the jobs is positive, which is rather surprising. However, an increase of 5 hours/week at work only increases the utility by as much as a 1% increase in the probability of graduating in 4 years. The coefficients on status of the jobs, hours/week spent on coursework, and earnings at 30 are not significantly different from zero.

Next I check for evidence of specialization. Estimates presented in Table B.10 suggest that there is strong evidence of extreme specialization for graduating in 4 years ($\triangle u_{12} > 0$, $\triangle u_{11} \approx 0$), and for finding a job ($\triangle u_{52}/\triangle u_{51} \gg 1$). This implies that individuals concentrate their chances of graduating in 4 years, and getting a job upon graduation in one of the majors in their major pair.[41] On the other hand, approval of parents and enjoying coursework are outcomes that are important in the choice of both majors (i.e. $\triangle u_{41} > 0$, $\triangle u_{42} \approx 0$ and $\triangle u_{31} > 0$, $\triangle u_{32} \approx 0$ respectively). The coefficient on hours/week spent on coursework, γ_{12}, is negative; this supports the specialization hypothesis, i.e. individuals prefer pairs of majors that entail different hours/week in college.

Columns (2) and (3) in Table B.10 shows the estimates for the males and females sub-samples respectively. The three most important outcomes for both are the same: enjoying the coursework, graduating in 4 years, and approval of parents (though not in the same order). For males, an analysis of the ratios of $\{\triangle u_{r2}/\triangle u_{r1}\}_{r=\{1,3,4,5\}}$, and γ_{12}/γ_{11} reveals that they prefer to choose majors that differ in their chances of graduating in 4 years ($\triangle u_{12} > 0$, $\triangle u_{11} \approx 0$), in enjoying the coursework ($\triangle u_{32}/\triangle u_{31} \gg 1$), and approval

[41]There is ample evidence of the latter in the comments submitted by the respondents (see Appendix 1).

of parents ($\triangle u_{52}/\triangle u_{51} \gg 1$). The coefficient on hours/week spent on coursework, γ_{12}, is negative implying that males prefer pairs of majors with different coursework levels. Females, like their male counterparts, prefer majors that entail different chances of graduating in 4 years ($\triangle u_{12} > 0$, $\triangle u_{11} \approx 0$). In addition, they prefer majors that differ in their chances of getting a job upon graduation ($\triangle u_{52} > 0$, $\triangle u_{51} \approx 0$). There is also some evidence of females preferring majors with different amounts of workload in terms of coursework ($\gamma_{12} < 0$). On the other hand, approval of parents and enjoying coursework matter significantly in the choice of both majors ($\triangle u_{41} > 0$, $\triangle u_{42} \approx 0$ and $\triangle u_{31} > 0$, $\triangle u_{32} \approx 0$).

This model exhibits the restrictive IIA property, which is not a very realistic assumption in this particular situation. For example, one could imagine that an individual majoring in Area Studies and Literature & Fine Arts is more likely to choose Area Studies and Ethics & Values, rather than Natural Sciences and Ethics & Values. To allow flexible substitution patterns, I allow for a stochastic part for each major that is perhaps correlated over majors and heteroskedastic over individuals and majors (these appear as 12 random effects, one for each of the 7 alternatives in WCAS, and the 5 categories outside WCAS), and another stochastic part that is iid over individuals and alternatives. The utility function of a major pair p is now:

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

where, as before, ε_{ip} is a random term with zero mean that is iid over alternatives of major pairs, and is normalized to set the scale of utility. The $\eta_{i,m}$ for $m = \{1,..,12\}$ are normally distributed effects with zero mean, and $c_{px} = 1$ if major x appears in the major pair p.[42] This structure allows flexible substitution patterns across alternatives. For example, the correlation between a major pair κ consisting of $m = \{1,2\}$, and a second major pair ω consisting of majors $m = \{2,3\}$ is $E([U_{i\kappa} + \varepsilon_{i\kappa} + \eta_{i,1} + \eta_{i,2}][U_{i\omega} + \varepsilon_{i\omega} + \eta_{i,2} + \eta_{i,3}]) = Var(\eta_{i,2})$. So utility is now correlated over alternatives. Given the vector $\boldsymbol{\eta}_i$, the *conditional* choice probability is simply logit, since the remaining error term is iid extreme value. The probability of individual i choosing the major pair p is:

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

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

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

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

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

The estimated parameters from maximizing the simulated log-likelihood, $\sum_i \ln(\widehat{P_i(\Omega)})$, are shown in Table B.12. The coefficients are similar in relative magnitude, but larger

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

in absolute terms than the corresponding fixed coefficients in column (1) of Table B.10. This is because, in the standard model, all stochastic terms are absorbed into one error term, ϵ. The variance of this error term is larger in the standard logit model than in a mixed logit since some of the variance is now captured by the η's rather than the ϵ in the mixed logit model. Since utility is scaled so that ϵ has the variance of an extreme value, the variance before scaling is larger in the standard logit than the mixed logit, and hence parameters are scaled down in a standard logit relative to the mixed logit. Graduating in 4 years, enjoying the coursework, and approval of parents continue to be the three most important outcomes. Individuals choose majors in their choice pair such that they enjoy coursework and have approval of parents in both majors ($\triangle u_{31} > 0$, $\triangle u_{32} \approx 0$ and $\triangle u_{41} > 0$, $\triangle u_{42} \approx 0$). Graduating in 4 years is an important consideration for both majors, but there is some evidence that individuals prefer majors that differ in their chances of graduating in 4 years ($\triangle u_{12}/\triangle u_{11} > 1$). Individuals also prefer pairs of majors that allow them different chances of getting a job upon graduation ($\triangle u_{52} > 0$, $\triangle u_{51} \approx 0$). Graduating with a GPA of at least 3.5 has a positive coefficient but is not significant. The somewhat puzzling results are the positive coefficients on hours/week spent on coursework, and at the jobs (the latter is not significant). γ_{12}, the coefficient on $\min[E_{ip_1}(d_1), E_{ip_2}(d_1)]$, is negative suggesting that individuals prefer pairs of majors with different time commitments at college. However, it is not significantly different from zero.

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

2.7. Understanding Gender Differences

The descriptive analysis in section 2.3.4 documents the heterogeneity in beliefs for various outcomes between the two genders. In sections 2.5 and 2.6, it is shown that males and females also differ in their preferences for the various outcomes. Though the results of the decomposition metric of equation (2.14) presented in Tables B.2, B.5, and B.8 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 past gender differences in beliefs about success in different occupations (see 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 were solely due to gender differences in preferences, then no direct policy intervention could change the gap. Alternatively, if the gender gap existed because of, say, gender differences in beliefs about ability and self-confidence, then policy interventions like single-sex classes could possibly reduce the gap.[43] In this section, I dig deeper into the underlying causes for the gender gap.

2.7.1. Decomposition Analysis

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

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

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

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

$$= [\overline{F(X_M\widehat{\beta}_M)} - \overline{F(X_F\widehat{\beta}_M)}] + [\overline{F(X_F\widehat{\beta}_M)} - \overline{F(X_F\widehat{\beta}_F)}]$$

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

where N_j is the sample size of gender j.[44] The first expression in the square brackets represents part of the gender gap that is due to gender differences in distributions of X, and the second expression represents the part due to differences in the group processes determining levels of Y. It is relatively simple to estimate the total contribution. However, identifying the contribution of group differences in specific variables/ coefficients to the gender gap is not straightforward. For this purpose, I use a decomposition method proposed by Fairlie (1999, and 2005). Contributions of a single variable/ coefficient are calculated by replacing the relevant variable of one group with that of the other group sequentially one by one. For illustration, suppose $Y_j = F(X_j \beta_j)$ for j=$\{F, M\}$, and that X includes two variables, X_1 and X_2. Moreover, let $N_M = N_F = N$, and assume there exists a natural one-to-one matching of female and male observations. The independent contribution of X_1 to the gender gap is given as:

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

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

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

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.[45] 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:

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

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

For the purposes of this decomposition, I treat double-major respondents as if they were pursuing a single major; I use the parameter estimates obtained from the single major choice model estimation using stated preferences of the respondents. Results of

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

this decomposition are presented in Table B.13 for four different majors.[46] The last row of the table shows that both expectations and preferences contribute to the gender gap for all major categories. The contributions of preferences and beliefs to the gap differ by fields: majority of the gender gap in Literature & Fine Arts and Social Sciences II is due to gender differences in beliefs, while gender differences in preferences explain most of the gap in Engineering and Social Sciences I.

A closer look at columns (1)-(4) shows that gender differences in beliefs about ability (more precisely beliefs about graduating in 4 years, and graduating with a GPA of at least 3.5) are insignificant and explain a small part of the gender gap. If women are less overconfident than men (Niederle et al., 2007; and references therein), and low in self-confidence (Long, 1986; Valian, 1998), one would expect females to have lower beliefs (relative to males) about graduating in 4 years and graduating with a GPA of at least 3.5, but that is not the case. Therefore, explanations entirely based on the assumption that women 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 B.13. 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

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

differences in beliefs about enjoying coursework. Interestingly, gender differences in beliefs about future earnings are insignificant and constitute less that 0.5% of the gap. Females have beliefs similar to those of males about academic ability in engineering.[47] These findings suggest that females are less likely to major in engineering not because they are underconfident about their academic ability, low in self-confidence, or because of beliefs about wage discrimination in the labor market. Instead this is because they believe that they won't enjoy taking courses in engineering. In other words, it's not that women think they won't be good engineers, but they think they won't enjoy studying it. The results seem to suggest that a policy that changes social attitudes might be more useful in narrowing the gap. In the next section, I study how the gender gap changes by simulating different environments.

2.7.2. Simulations

I carry out some simulations to see how the gender gap would change in a world with a different environment. Column (1) of Table B.14 shows the gender gap predicted by the model for the various major categories. The simulation in column (2) considers an environment where the female subjective ability distribution (beliefs about graduating within 4 years, and about graduating with a GPA of at least 3.5) is replaced with that of males.[48] The purpose of this simulation is to answer how much of the gap is due to females having less self-confidence in their ability. The second simulation in column (3)

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

[48] 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.

replaces the female subjective earnings distribution with that of males; it is meant to answer the question of how much of the gap is due to beliefs of wage discrimination in the labor market. Columns (4) and (5) simulate an environment in which females have the same beliefs as males about enjoying coursework and enjoying work at potential jobs respectively.

I continue to focus the discussion on Engineering. The results confirm the findings obtained in Table B.13. If female expectations about ability were raised to the same level as that of males through some policy intervention, the gender gap in engineering would decrease by less than 14%. The gender gap virtually stays the same if female expectations of future earnings were forced to be the same as those of males. Finally, the gender gap reduces 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. It is not clear what kind of policy would be able to bring about a change in the female beliefs about enjoying coursework. This is because gender differences in beliefs of enjoying coursework are hard to explain: they could be a consequence of innate gender differences in attitudes (Baron-Cohen, 2003), or due to social biases including discrimination (Etzkowitz et al., 1992; Valian, 1998).[49] However, the insignificant and small gender differences in ability and future earnings in engineering allows me to rule out low self-confidence in women and perceived wage discrimination in

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

the labor market as possible explanations for why women are less likely to major in fields like engineering.

A major question that has been left unanswered is the source of gender differences in preferences. Gender differences in preferences could arise from differences in tastes, as well as gender discrimination. For example, parents who know that females would be discriminated in male-dominated majors/ occupations could try to shape the preferences of their female children so that they are more comfortable in female-dominated majors/ occupations (Altonji and Blank, 1999). The question of understanding the sources of gender differences in preferences is beyond the scope of this paper.

2.8. Conclusion

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

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

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

The analysis in this paper has some limitations. First, the study is based on data from Northwestern only. The heterogeneity in subjective expectations underscores the need to elicit similar data at different undergraduate institutions, and at a larger scale in order to make policy recommendations. Second, heterogeneity in subjective responses could be driven by differential access to information, or by different information processing. Demographic data collected from respondents allows me to explain some of the heterogeneity in beliefs; I find that cultural proxies and parents shape beliefs for certain outcomes. However, progress in understanding how people form and update expectations requires richer longitudinal data. Moreover, as Manski (2004) argues, understanding expectations formation will also require intensive probing of individuals to learn how they perceive environments and how they process new information. Third, individuals may find it optimal to experiment with different majors to learn about one's ability and match quality (Manski, 1989; Altonji, 1993; and Malamud, 2006). This study does not focus on this aspect by assuming that individuals maximize current expected utility. Since experimentation may be important, I plan to focus on it in follow-up work.

My results shed some light on the reasons for the gender gap in college major choice. Gender differences in beliefs about ability and future earnings are insignificant in explaining the gender gap. A policy intervention which were to raise the expectations of females about ability and future earnings in engineering to the same level as that of males would only decrease the gender gap by about 15%. This has two implications: (1) just raising expectations of women may not be enough to eradicate the gap, and (2) hypotheses which claim that the gap could be explained by women having low self-esteem and being less overconfident than men can be rejected by my data. Most of the gender gap is due to gender differences in beliefs about enjoying coursework, and preferences for various outcomes. The evidence suggests that social prejudices and wage discrimination may not be the main explanation for why women are less likely to major in engineering. However, one should be careful in jumping to a definite conclusion since gender differences in beliefs about enjoying coursework as well as preferences may exist because of differences in tastes, or due to gender discrimination. Richer data is needed to answer this question. I believe the next natural step is to re-interview respondents in my sample to explore these issues.

CHAPTER 3

Social Conformity: Theory and Experimental Investigation

"We do not live exactly as our parents lived but whatever we do now is only a modification of what was done before. It could hardly be otherwise. Very little of our public behavior is innate; most of us have only very limited creative originality. We act as we do because, one way or another, we have learned from others that this is the way we ought to behave." (Leach 1982: 128)

3.1. Introduction

A positive correlation between an individual's choice of college major with that of his reference group is consistent with either the individual (1) learning about that particular choice through the experiences of others, and hence choosing that major (social learning), (2) getting a utility gain by simply having the same major as one's reference group (social comparison), or (3) sticking to the norm because of image-related concerns (social influence).[1] This distinction is important for at least two reasons. First, it is relevant for our theoretical understanding of the specific processes through which individual choices are made. Second, it has important policy implications. In the example above, conformity arising from social learning may be desirable since it leads to more informed choices. Social interactions have been an active area of economic research for some time now with studies

[1] A fourth possible explanation is the genetic transmission of preferences and beliefs (Bisin and Topa, 2003). However, this is not the focus of the current study.

focusing on a wide range of empirical settings including teenager experiments with illegal drugs (Duncan et al., 2005), conforming to the behavior of peers at high school (Cooley, 2006) and at college (Sacerdote, 2001), and coordinating fertility practices (Kohler et al., 2001; Munshi et al., 2006). However, most studies focus on measuring the extent of social interactions, and very little attention has been given to studying the mechanisms through which they are generated. In this paper, I focus on disentangling the role of social learning and social comparison from that of social influence in social interactions.

I develop a simple model constructed on the premise that people are motivated by their own payoff, and by how their action compares to others in their reference group. Individuals compare their actions to choices of others because they either believe that the choices of others provides a stronger indication as to what the correct course of action is (as in the case of Banerjee, 1992, and Bikhchandani et al., 1992), they get a utility gain by making the same choice as their peers even when there is no uncertainty about the *intrinsic* utility maximizing choice (Cialdini, 1993), or they want to avoid the discomfort of being different from others (as in Asch, 1958).

The problem is that learning about the norm (either through the channel of social learning or social comparison) and image-related concerns may both generate the same empirical facts. In order to disentangle the two channels empirically, I exploit the fact that conformity arises from image-related concerns only if the individual's actions are observable to other people. To this end, I design and conduct an experimental investigation of a charitable contribution game that unmasks subjects' choices in a systematic and controlled way. Even though an individual is assigned to a group, his payoff only depends on his own actions, i.e. there are no externalities in the compensation scheme across group members. The experimental setting, besides providing an environment that provides clean evidence on each of these mechanisms, also allows me to overcome the difficult identification problems in measuring social interactions in real-world settings (Manski, 1993, 2000; Moffitt, 2001).

The empirical results show that both learning about the norm (social learning or comparison) and social influence lead to conformity in actions. In the absence of any social information, actions of individuals are correlated with their beliefs about the group action. Conformity in actions arises when individuals are informed about some statistic of the contribution of other group members. The experimental design is able to disentangle social influence from other mechanisms as a channel that leads to conformity. However, it does not allow me to distinguish between social comparison and social learning.

The second set of results deals with the role of social ties in my experimental setting. I am not aware of any experimental study that analyses the role of social ties in settings where there are no externalities arising from the monetary incentive scheme across individuals. Bandiera et al. (2007) find that an intermediate norm evolves in the presence of friends in a field study where each individual's payoff only depends on his own actions. Once exact group membership is observable to an individual, I find that the effect of social influence varies by the nature of social ties (defined by which group members the individual already knows from outside the lab) in the group. In particular, individuals only change their actions in response to the choices of group members they are *friends* with. Moreover, a low contribution norm evolves that causes individuals to contribute less in the presence of friends.

This paper is organized as follows. In section 3.2, I start out with a brief literature review and outline several channels of conformity. Section 3.3 describes the theoretical model in which conformity arises from learning about the norm and social influence. In section 3.4, I describe the experimental setup which is used to test the model predictions. Section 3.5 presents the empirical analysis. Finally, section 3.6 concludes.

3.2. Literature Review

The classic work on conformity is the experiment conducted by Asch (1946). Subjects were placed in groups whose other members were secretly confederates of the researcher; they were asked to estimate the geometric length of a line by matching it with one of three lines after some of the other group members had given their opinion one at a time. In cases when the confederates unanimously endorsed a clearly wrong comparison line, about one-third of the tested subjects conformed to the wrong judgement of the false majority. Individuals in the control group (not under social pressure) answered correctly with a few exceptions. Since then, social psychologists have developed several theories of conformity (see Deutsch and Gerard, 1955; Cialdini, 1993).

One of the earlier theoretical economic works on conformity is Jones (1984). He presents a model of exogenous conformism in which a penalty is added to the utility function that depends on the distance between the individual's choice and that of all the other group members. His formulation of penalty induces the individual to shift his choices towards the average. His analysis is restricted to the case of two individuals, which does not seem to be adequate when addressing the issue of conformity. Bernheim (1994) derives conformity endogenously. His main assumption is that individuals, in addition to consumption, care about status, which is inferred from their actions. He shows uniformity around a value that is exogenously regarded as the best by agents.

Economists have increasingly been interested in empirically investigating "peer effects". However, the goal of these studies has mostly been to measure the strength of these social interactions, and very little is known about why conformity may arise. Depending on the context, conformity may arise through the following channels: (1) social learning, (2) social comparison, (3) strategic complementarities, and (4) social influence.

Conformity through social learning may arise if one's private signal/ information is not a sufficient statistic, and more information about the correct action may be learned through the choices of others. An example of this is the investigation of the role of social learning in the diffusion of a new agricultural technology in Ghana (Conley and Udry, 2005). There is extensive theoretical literature on conformity arising because of herding (Banerjee, 1992) and information cascades (Bikhchandani et al., 1992).[2] Conformity through social learning will arise even if the individual's identity stays private, but requires that there be uncertainty about the utility maximizing action.

Social comparison may lead to conformity if agents use the behavior of others as a reference point for decisions (Cialdini, 1993; Cason and Mui, 1997; Messick, 1999). This would arise even in settings where there is no uncertainty about the utility maximizing action.

Conformity may also arise because of strategic complementarities. An example of this is Sweeting (2006) who shows that radio stations coordinate the timing of commercial breaks so that fewer listeners avoid commercials and the value of advertising time is

[2] Also see Anderson and Holt (1997), and Goeree and Yariv (2006) for their experimental investigation.

increased. This channel has also been investigated in experimental public provision good games in which subjects receive information about others' choices (for example, Brandts et al., 2001; Keser and van Winden, 2000). In these experiments, a subject's payoff depends on both his choice as well as the choice of others. In the presence of this strategic interdependence, it is hard to discern how much of the change in a subject's behavior (after he observes others' choices) is due to social interactions, and how much is due to the subject's attempt to increase his own payoff. Empirically very little is known about why agents conform unless there are strategic complementarities. One exception is the recent empirical investigation of productivity of checkers for a grocery store by Mas and Moretti (2006). They find strong evidence of positive productivity spillovers from the introduction of highly productive personnel into a shift. More interestingly, they find that a worker's effort is positively related to the presence and speed of workers who physically face him, but not the presence and speed of workers whom he faces (and who do not face him). This implies that workers do not like it when faster colleagues are looking at them, either because they fear being accused of slacking off, or because they feel inferior or stigmatized even without accusation. Moreover, workers respond more to the presence of co-workers with whom they frequently interact. These patterns suggest that image-related concerns may play an important role in explaining conformity.

This leads to the final explanation for conformity- social influence. Individuals may conform to the choices of others if they fear some form of social sanction if they were to deviate from the social norm. This norm-based approach for conformity has been used to explain the fertility transition in Bangladesh by Munshi et al. (2006). Social influence may only lead to conformity if both the individual's identity and actions are observable to his reference group. The role of revealing identity has also been studied in public good experiments; Andreoni et al. (2004) and Rege et al. (2004) find that revealing one's identity increases the contribution significantly in public good games.

In real-world instances, measuring social interaction effects raises difficult identification problems (see a detailed discussion in Section 3.4). Lately, field and laboratory experiments have been employed to study social interactions. Experimental economists have attempted to explain the positive correlation between an agent's action and the social choice through theories of reciprocity (Rabin, 1993). For reciprocal motives, it must be the case that others' behavior matters through its effect on the individual's welfare. This is true in the case of public goods, but is implausible for various examples of social interaction effects. Conformism differs from reciprocity since reciprocal behavior depends on the welfare effects of the stimulus behavior, while conformist behavior does not. Consider a simple example - a conformist will contribute to a useless public good (which benefits no one) if he observes others making contributions, but a reciprocity-motivated agent will not since he does not benefit from the behavior of others.

Conditional cooperation has been studied in some other contexts as well. Cason et. al. (1998) investigate social influence in a sequential dictator game. They conclude that subjects become more self-regarding (selfish in the sense of maximizing the allocation that goes to them) in the Irrelevant Information treatment, and that observing relevant information constrains some subjects from moving toward more self-regarding choices. The problem with their setup is that very little information is imparted to the subjects. To observe social influence, more information about normal behavior needs to be imparted.

Conditional cooperation has been studied in field and lab experiments in charitable contributions as well (see, for example, Frey and Meier, 2004; Landry et al., 2006; Croson and Shang, 2005). Again the problem is that the results are consistent with at least two theoretical approaches: people may want to conform to a social norm, or, contributions by others may serve as a signal of the quality of the charity.

This paper tries to disentangle the role of social learning and social comparison (which turn out to be indistinguishable in my experiment) from that of social influence. Due to the reasons mentioned above, I specifically consider an experimental setup in which there are no payoff complementarities across individuals. I use the term "leaning about norm" to imply both social learning and social comparison.

3.3. Model of Charity Contribution

This section outlines a simple model of charitable contribution under different environments. The goal is to come up with predictions that can be tested in an empirical setting.

3.3.1. No Information Case

Consider an environment consisting of many agents, each of whom has an endowment normalized to 1. The agent selects $x \in [0, 1]$ which is given to a charity. The agent has intrinsic preferences over what he keeps $(1 - x)$ and, depending on his type, may have preferences over what is given to the charity. Moreover, the choice is only observed by the agent himself, and is not publicly observable. A *self-interested* agent will keep his entire endowment for himself since he gets no utility from contributing to the charity.

On the other hand, an individual may donate to a charity because of pure altruism or warm-glow. A pure *altruist* cares about the payoff to himself as well as the total payoff to the charity, while an individual motivated by *warm-glow* gets utility from the mere act of contributing.[3,4] I define the different types of agents as follows.

Definition 1. *An agent i is self-interested if his utility function has the form $u(1-x_i)$, is induced by altruism if his utility function looks like $u(1-x_i, \sum_{j=1}^{N_i} x_j)$, and is motivated by warm-glow if the utility function has the form $u(1-x_i, x_i)$.*

The utility function for an agent is twice continuously differentiable, and strictly increasing and concave its arguments.[5] An altruistic agent cares about the total contribution to the charity, and, in the absence of any information about others' contributions, has beliefs $P(\sum_{j=1}^{N_i} x_j)$ about the contributions of others. An altruistic agent chooses $x_i^* \equiv \arg\max_{x_i \in [0,1]}(\int u(1-x_i, \sum_{j=1}^{N_i} x_j) dP_i(\sum_{j=1}^{N_i} x_j))$. A self-interested agent will contribute zero. Depending on the exact functional form of the utility function, an agent motivated by warm-glow will donate $x_i \in (0,1]$, and an agent motivated by altruism will donate $x_i \in [0,1]$. Let x_i^* be the choice that maximizes the agent's intrinsic (expected) utility function

[3]See Andreoni (1990) for a theoretical model of warm-glow giving, and Anderson et al. (1998) for a theoretical analysis of altruism.

[4]There could be a fourth kind of agent- an impure altruist (Andreoni, 1990). His utility function would be of the form $u(1-x_i, \sum_{j=1}^{N_i} x_j, x_i)$. A purely altruistic agent and an agent motivated by warm-glow are special cases of this. Since the decision problem of an impure altruist is similar to that of a pure altruist, I don't consider this case separately.

[5]I am somwhat sloppy with the notation in the sense that the function $u(.)$ in $u(1-x_i)$, $u(1-x_i, x_i)$, $u(1-x_i, \sum_{j=1}^{N_i} x_j, x_i)$ is actually a different function in each of these cases.

in this case. Based on this definition, observing the contribution of an agent does not allow the researcher to infer his type.[6]

3.3.2. Limited Information Case

Next I look at how contributions of the different types of agents are affected by receiving some limited information, for example, the group average of the previous round contributions. The key feature of the limited information environment is that the identities of the group members remain private, and so prestige concerns are absent. Information about the contribution of others may serve as a signal of the quality of the charity (Vesterlund, 2003); I eliminate this channel by only considering the case of a well-known charity like the Red Cross. In the absence of concerns for social comparison, receipt of some information about others' contributions should only alter the contribution of an altruistic agent via the updating of his beliefs, $P(.)$.

Revelation of some additional information may alter the contributions of an agent if he engages in social comparison by using the behavior of others as a reference point for decisions (Cialdini, 1993; Cason and Mui, 1997; Messick, 1999). Therefore, conformity may arise in the limited information case either because of (1) social learning (for an altruistic agent), or (2) social comparison (for any type of agent). Since empirically it will not be possible to distinguish between the different types of agents, I only focus on social comparison concerns in the model outlined in this section. The utility function is now modified to include a loss term that depends on some distance metric between the agent i's choice, x_i, and the group choice, s_{-i}. The group choice is simply defined as the

[6]An agent contributing zero could be one who is selfish or motivated by pure altruism; an agent contributing a positive amount could be one motivated by altruism or warm-glow.

statistic of others' contribution which is provided to the agent.[7] Each agent differs in the importance he places on how his choice compares to the social norm, and this is indicated by the parameter $t \in [0, \bar{t}]$. The value of t is an agent's private information. The utility function for agent i is now:

$$U(x_i, t_i) = \mathbf{1}_{SI} * u(1 - x_i) + \mathbf{1}_A * u(1 - x_i, \sum_{j=1}^{N_i} x_j) + \mathbf{1}_{WG} * u(1 - x_i, x_i) + t_i G(x_i, s_{-i})$$

where $\mathbf{1}_{SI}$, $\mathbf{1}_A$, and $\mathbf{1}_{WG}$ is an indicator function that equals one if the agent is self-interested, altruistic, or motivated by warm-glow respectively. $G(x_i, s_{-i})$ is the penalty function for not conforming to the group choice, s_{-i}. Agent i cares about his own choice x_i, and may care about how his choice compares with the group choice s_{-i}. The parameter t captures the extent to which the agent cares about social comparison.

The size of the penalty function depends both on the form of the penalty function $G(.)$ and the form of the loss function relating x_i and the group choice, s_{-i}. I define the following metric to capture the distance between the agent's choice and the group choice:

$$\frac{|x_i - s_{-i}|}{\sigma_{-i}}$$

where x_i is individual i's choice, and s_{-i} is some statistic of the distribution of choices of the other group members. The distance from the group choice is normalized by the standard deviation in the choices of other agents, σ_{-i}, to capture the fact that when choices have a large dispersion, individual i feels less pressure to conform (Messick et al.,

[7]In cases where the agent is informed of the entire contribution distribution of other group members, it is not clear what statistic (mode, median, average etc.) the agent uses as the group choice. I test various statistics in the empirical section.

1983). The utility function is now:

$$U(x_i, t_i) = \mathbf{1}_{SI} * u(1 - x_i) + \mathbf{1}_A * u(1 - x_i, \sum_{j=1}^{N_i} x_j) + \mathbf{1}_{WG} * u(1 - x_i, x_i) + t_i G(\frac{|x_i - s_{-i}|}{\sigma_{-i}})$$

The penalty function $G(.)$ is continuous, strictly concave, twice differentiable, and $\arg\max G(.) = 0$ ($G'(s) < 0$ and $G''(s) < 0$ for $s > 0$).

I first consider the case where social comparison concerns are absent, i.e. $t_i = 0$.

Claim 1. *In the absence of social comparison, upon receipt of some information about others' contributions, the contribution of a self-interested agent and that of an agent motivated by warm-glow will be unaffected, while that of an altruistic agent may change (relative to the no information case).*

Proof. In the absence of social comparison $t_i = 0$. The utility functions of a self-interested agent and an agent motivated by warm glow are functions of own contributions only; their contributions will stay unaffected since $\arg\max U(x_i, t_i) = x_i^*$. Conversely, even in the absence of concerns for social comparison, an agent motivated by altruism cares about the total amount contributed to the charity, and hence the contribution of others shows up in his utility function; depending on the specification of the utility function, receipt of information about the contribution behavior of the group may cause him to change his own contribution. □

Now consider the case where concern for the social norm may be present. Given the concavity assumptions, the best choice is unique. Let $x^{**}(t)$ denote the private optimum.

Proposition 1. *For* $\forall t$, $x^{**}(t) \in [0, s)$ *for a self-interested agent, and* $x^{**}(t) \in [0, 1]$ *for an agent motivated by altruism or warm-glow.*

Proof. See Appendix C.4. □

This proposition states that, in the presence of social comparison concerns, a self-interested agent would contribute less than the group choice. However, an agent motivated by altruism or warm glow could contribute an amount in the entire range. I next show some comparative statics.

Proposition 2. *(a) In the case of a self-interested agent (or an agent motivated by altruism or warm-glow and* $x^{**}(t) < s$*),* $x^{**}(t)$ *is:*

(1) weakly increasing in t

(2) increasing in s,

(3) decreasing in the dispersion, σ, *of the contributions of others.*

(b) In the case that $x^{**}(t) \geq s$, *and the agent is motivated by altruism or warm-glow,* $x^{**}(t)$ *is:*

(1) weakly decreasing in t

(2) increasing in s,

(3) increasing in the dispersion, σ, *of the contributions of others.*

Proof. See Appendix C.4. □

The results are intuitive. In a regime where $x^{**}(t) < s$, higher types (defined by a higher t) choose a higher value of x. The choice is increasing in the group choice, and decreasing in the dispersion of others' contributions. This is a consequence of the way the conformity index was defined, since the degree of conformity decreases if the choices of other agents are more dispersed. In a regime of full conformity, $x^{**}(t) = s$, increases in s and σ have no effect since the conformity term no longer matters.

From proposition 1, we know that $x^{**}(t) \geq s$ for an altruistic agent or one motivated by warm-glow. In the regime $x^{**}(t) \geq s$, a higher type (one who cares more about the social norm) will choose a lower value of x in order to be closer to s. The choice is increasing in the group choice, and in the dispersion of the contributions. The following claims follow directly from Propositions 2:

Claim 2. *In the limited information case, because of social comparison concerns, contributions should move closer to the group choice, s.*

Claim 3. *In the limited information case, contributions are increasing in the group choice.*

3.3.3. Full Information Case

I now consider the full information case, i.e. an environment where the individual receives information about choices of group members as well as their identities. Under a full information case, incentives that can be important for contributing are prestige (Harbaugh, 1998a, 1998b), social comparison, social approval, and avoiding shame (Elster, 1999). For prestige to be a motivation, identification of one's contribution by other group members is necessary. Individuals motivated by social comparison care about how their contribution compares to that of the other group members. Individuals motivated by social approval (or avoiding shame) are concerned about how their contributions will be perceived by other group members.[8] I now modify the utility function to include the effect of image-related concerns and social approval. Each agent now also cares about his image, p. More specifically, the agent cares about how his type is perceived by other agents. Thus, the utility function is:

$$U(x_i, t_i, p_i) =$$
$$\mathbf{1}_{SI} * u(1 - x_i, p_i) + \mathbf{1}_A * u(1 - x_i, \sum_{j=1}^{N_i} x_j, p_i) + \mathbf{1}_{WG} * u(1 - x_i, x_i, p_i) + t_i G(\tfrac{|x_i - s_{-i}|}{\sigma_{-i}})$$

I further assume the utility is also twice differentiable, strictly increasing and concave in image, p. I first consider the case when types are observable. In that case, perception p of a type t agent equals his type, i.e., $p = t$. Given the concavity assumptions, the private optimum is unique. Let $\widetilde{x(t)}$ denote the optimal choice. Moreover, assume that consumption and image are weak substitutes (this requires the assumption that $u_{12} \leq 0$ for a self-interested agent, and $u_{13} \leq 0$ for an agent who is motivated by altruism or warm-glow), and that contribution and image are compliments for an agent motivated by altruism or warm-glow ($u_{23} > 0$). Under these assumptions, in the full information case, a type t agent will contribute at least as much as in the limited information case. This result is a consequence of the fact that image and consumption are now substitutes: the individual is now willing to forgo some consumption to build a more favorable perception. This is stated formally in the following claim:

[8]Social approval does not mean that the agent is concerned about other people knowing how much he contributes, but instead that the agent is concerned about how other people evaluate his contribution.

Claim 4. *The cumulative distribution of contributions in the full information case first-order stochastically dominates the cumulative distribution function of contributions in the limited information case.*

Proof. See Appendix C.4. □

Finally, there will always be some agents who will contribute zero, or their entire endowment. The following claim tells us that there is always mass at zero and at 1 in the limited information case.

Claim 5. $\exists\, t^* > 0$, *such that* $x^{**}(t) = 0$ *for* $t \leq t^*$. *Also* $\exists\, \widehat{t} > 0$, *such that* $x^{**}(t) = 1$ *for* $t \geq \widehat{t}$

Proof. See Appendix C.4. □

The result will also hold in the full information case.

The claims that have been made so far assume that the type of an agent is observable. However, this is generally not the case. In the case where types are unobservable, the beliefs of other agents about $i's$ type is denoted by Φ. On seeing i's choice of x, they form an inference about i's type which is given by $\Phi(x)$. Agent i does not see the inference directly, but knows the equilibrium relation between x and t, and takes this into account. Therefore, this is a signalling game. The signalling equilibrium consists of a choice function (mapping types to choices), $\mu : t \rightarrow x$, and an inference function (mapping choice to inferences), $\Phi : x \rightarrow p$, where choices are optimal given inferences, and inferences are consistent with choices. I next show that the choice function is weakly monotonic.

Proposition 3. *In any signalling equilibrium, if $t > t'$, then $\mu(t) \geq \mu(t')$. That is, more conforming types will make choices that are weakly more conforming to the social norm.*

Proof. See Appendix C.4. □

In principle, one can show the existence of both a separating equilibrium, and a pooling equilibrium under certain conditions.[9] However, the goal of this section was to come up with testable hypotheses for the empirical setting. Having done that, I move to an empirical setting which will allow me to test the various claims outlined in this section.

3.4. Experiment

In real-world instances, measuring social interaction effects raises difficult identification problems because interdependent behavior can take different forms that are difficult to isolate. In Manski's terminology (1993), an individual's behavior may vary according to the endogenous behavior of the group, but it may also vary with the exogenous characteristics of the group members. Moreover, outcomes need not arise from interdependent behavior: members of a given group may behave similarly because they have similar unobserved characteristics or face similar institutional environments (correlated effects). In a simple linear-in-means model, Manski (1993) shows that equilibrium outcomes cannot distinguish endogenous effects from exogenous effects or correlated effects. In this context, it is impossible to identify the true nature of social interactions. Even if one were to overcome these identification issues, estimation raises serious econometric problems. The

[9] The model in the full information case is similar (in some ways) to the framework in Andreoni and Bernheim (2007). They explain the norm of equal splits in a dictator game as a desire to be perceived as fair, and model this as a signalling game.

mean group behavior (which appears as a regressor) can be endogenous for the following reasons: If individuals self-select within groups, they are likely to face common shocks and their unobserved characteristics are likely to be highly correlated (sorting bias). Moreover, in small groups, individual and group behavior feed on one another, and thus they are potentially simultaneously determined (simultaneity bias).

Therefore, I consider a laboratory experiment setting, as laboratory experiments have many advantages over alternative sources of information for the purpose of estimating social interactions. Randomization of participants across groups limits correlated effects and sorting biases. Experiments allow one to control the reference group with whom individuals interact in the laboratory. Moreover, group size can be determined exogenously and membership assigned randomly. This clearly helps identify the endogenous and exogenous interactions effects. The main shortcoming of laboratory experiments is that they may lack external validity.

3.4.1. Experimental Design

The setup outlined for the theoretical model in section 3.3 is implemented as a charitable contribution game. Each subject is assigned to a reference group, members of whom are initially unknown to the subject. The experiment consists of six rounds.

The stage game in round 1 is as follows: Each subject is endowed with $10 dollars, and has the option to make a contribution to the American Red Cross. A well-known charity is picked to rule out the possibility that the contributions of others serve as a signal of the quality of the charity. The subject has to pick a contribution, x, where x can be any number between 0 and 10 in multiples of $0.5. If the subject decides to contribute to the

Red Cross, there is a 40% chance that the contribution does not go through, in which case the Red Cross gets nothing and the subject keeps his $10. Thus, there are two ways in which Red Cross does not receive anything: either the subject decides not to donate anything, or he decides to donate but the donation does not go through. The reason for this additional complication is as follows: in a round where only the amount which the Red Cross receives from a subject is made public, an agent may continue not to donate since other group members will not know whether his donation did not go through, or whether he did not donate in the first place. This modification allows an extra layer of masking. The instructions given to the subjects at the beginning of the experiment are in Appendix C.1.

The stage game is the same in each of the rounds. However, each round differs from the previous one is some way. More specifically:

- Round 2 is the same as Round 1 except that each subject is informed of the average amount Red Cross received from her group in Round 1 before he makes a decision.
- In Round 3, the subject is informed of the average amount Red Cross received from his group in Round 2, and is then asked to make contribution decision with the knowledge that the amount Red Cross receives from him in this round will be made public to other group members in Round 4. However, the subject is informed that identities will not be revealed.
- In Round 4, the subject observes the entire contribution distribution of his group (i.e. the amount Red Cross receives from each group member) before making the contribution decision. He is informed that the amount which Red Cross will

receive from him in this round and his identity will be made public to other group members in Round 5.

- In Round 5, subjects observe the amount Red Cross received from each group member in Round 4 along with their identities. They are then asked to make the same contribution decision. In addition they are told that other group members will observe their identity, the amount Red Cross receives from them, and their exact action (donate; not donate) in the next round. This round removes all uncertainties.
- Round 6 is analogous to Round 5.

The idea behind this design is as follows: Round 1 corresponds to the no information case. Round 2 corresponds to the limited information case: the subject makes the choice after observing the group choice (in this case the average of other group members' choices). Round 3 is similar to Round 2 in the sense that only information about group choice is given (and identities are not revealed). If a subject changes his contribution in Round 2 (relative to Round 1) or in Round 3 (relative to Round 2) after learning about the average donation from the previous round, it could either be attributed to social comparison, or an altruistic subject changing his contribution because he updates his beliefs about what the charity receives from everyone else. Round 4 is the first round in which the subject has to make a choice knowing that his identity will also be revealed along with his contribution. The change in contributions in this round relative to the third round is attributable to concerns for prestige, social approval, and avoiding shame. Round 5 only differs from the fourth round in that all information is made public i.e. all layers of masking are removed. Finally, round 6 is a repetition of the fifth round.

In Round 1, I also elicit subjective beliefs from subjects about the average donation of other group members. In order to incentivize the subjects to report their actual beliefs, a monetary reward is awarded if their guess is within a certain range of the actual group average. At the end of the session, data on some demographic characteristics was collected from the subjects. In addition, subjects were also administered the Marlowe-Crowne test M-C 2 (10). The test consists of ten questions concerning personal attitudes to which the subject has to respond either yes or no. The responses are then matched with a scoring algorithm, and the final score ranges from zero to ten, with a higher score corresponding to higher social desirability.[10]

3.4.2. Experimental Procedure

Subjects were recruited by posting flyers around campus and on Facebook, and by E-mailing the Northwestern economics undergraduate listserv. The study was advertised as an online economic experiment of decision-making. Subjects were informed that the study would last at most half an hour, and that they could earn as much as $15. A total of 101 subjects were recruited: 55 of them were females; 42 were majoring in Economics. Nearly half of the subjects were freshmen or sophomores.

The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) in the Northwestern Main Library Computer Lab. Nine sessions were held in total. Subjects were assigned randomly to a group. Across all sessions, there were 11 groups of 4 subjects each, 3 groups of 5 subjects each, and 7 groups of 6 subjects each. Each

[10]See Mandell for a discussion of the shorter version of the M-C Social Desirability Scale, and Appendix C.2 for the questions on the test.

session had at least two groups playing the experiment simultaneously. This was done so that individuals could not tell with certainty who the other members in their group were.

The following method was used to reveal the identities of the group members: each subject got to see the names of his group members on his screen at the start of round 5. In addition, the location of each group member in the room was shown as a matrix on the whiteboard in the front of the class at the start of the fifth round.

Students were paid $4 as a show-up fee. In each round, they were endowed with $10 and had to decide how much to donate to the Red Cross under different settings. At the end of the experiment, one round was chosen at random to determine the payoff to the subject and the Red Cross. This was done so that subjects had an incentive to treat each round as if it were a real round. Subjects were made aware of this feature of the experiment. Subjects earned an average of $12.67 (standard deviation of $2.83).

Subjects were provided with a hardcopy of the instructions which were also read aloud at the beginning of the experiment. In addition, a handout explaining the different projects undertaken by the Red Cross was provided to the subjects (see Appendix C.1). They were informed that, if they decided to donate, they could direct their donation to any cause of their choice from the list.[11] Donations were submitted to the Red Cross only after the end of the session; receipts of the donations to the Red Cross were sent by E-mail to all donors. The experiment concluded with debriefing the subjects.

[11] The options were: National Disaster Relief Fund; International Response Fund; Your Local Red Cross Chapter; Military Services; Measles Initiative; Blood Services Campaign.

3.5. Empirical Analysis

3.5.1. Experimental Results

Table C.1 presents some statistics for each round. The first two columns of the Table show that Claim 5 (presence of subjects who donate none or all of their endowment) holds in the data: in every round, there are at least 7 subjects who contribute their entire endowment, and at least 40 subjects who don't contribute anything.

Before I start the round-specific analysis, I test if the contribution pattern between any two consecutive rounds is different or not. This is done to assess the effect of the change in the stage setup on contribution behavior. Since there might be dependence between the contribution behavior of an individual across rounds, I use the (non-parametric) Wilcoxon signed rank sum test to check if there is a treatment effect between any two consecutive rounds. According to the results of the test presented in Table C.2, a significant (at the 5% level) and positive effect is found on the contributions of Round 4 which is the first round in which subjects make a decision knowing that their identities will be revealed.

In order to understand the contribution behavior of the subjects, I analyze one round at a time. In round 1, the subject makes a decision without any group information. As mentioned in section 3.3, one cannot pin down the type of an agent based on his contribution. For example, the 45 subjects who don't contribute anything to the charity in the first round could either be self-interested or motivated by altruism. I check if there is a positive correlation between the individual's contribution in Round 1 and his (elicited) expectation of the group average. If an individual believes that the average action of others is the *appropriate* behavior in this context (social proof; Cialdini, 1993), or if agents are conditional cooperators (as in Fischbacher et al., 2001), then one would expect to find a positive correlation between one's own contribution and his beliefs about what others will be doing. I, therefore, test the hypothesis:

H_1 : *Round 1 contribution* (x_1) *is correlated with the agent's expectation of contributions of other group members* ($guess1$)*.*

Before measuring the *degree* of correlation between x_1 and *guess1*, I test for the *existence* of correlation between x_1 and *guess1* by using the Spearman rank correlation coefficient (r_s). I find that $r_s = 0.50$, and that H_1 cannot be rejected at the 1% level. I interpret this as support for the hypothesis H_1, and then measure the degree of correlation. There is a positive correlation of 0.45 between x_1 and *guess1*.

Round 2 corresponds to the limited information case. The subject is informed of the average contribution of the group from the previous round (*avg1*). Proposition 1 states that all self-interested subjects would donate an amount $\in [0, s_{-i})$. Analysis of data shows that only 28 subjects donate an amount equal to or greater than the first period average. Therefore, there are at least 28 subjects across the sessions who can be classified as being motivated by altruism or warm-glow. I next check if claim 2, i.e., that the contributions of subjects should move closer (in this round) to the revealed group choice, holds in my data. This leads to the following hypothesis:

H_2 : *Relative to the Round 1 contributions* (x_1)*, Round 2 contributions* (x_2) *are closer to the group choice* ($s_{i,2}$)*.*

I define $Diff_{i,t} \equiv x_{i,t+1} - x_{i,t}$, i.e. $Diff_{i,t}$ is the change in subject i's contribution from round t to round $t+1$, and $M_{it} = s_{i,t+1} - x_{i,t}$, where $s_{i,t+1}$ is the group choice that is presented to subject i in period $t+1$ before he chooses his $x_{i,t+1}$. In the second period, the group choice ($s_{i,2}$) is the period 1 average contribution of group members. To test H_2, I consider the following regression:

$$Diff_{i,t} = \alpha + \beta M_{it} + \varepsilon_i \tag{3.1}$$

for $t = 1$. Here $M_{i1} = s_{i,2} - x_{i,1}$, i.e. it is the distance between the average round 1 contribution of the group and the agent's contribution in round 1. Support for H_2 would require $\beta > 0$. Column (1) of Table C.3 shows that this is indeed the case (standard errors are corrected for clustering at the group level in all the regressions). The estimate of β is positive and significant: the estimate of 0.31 implies that for a one dollar difference between the group choice and the subject's contribution, the subject's contribution moves, on average, by 31 cents in the direction of the group choice. Since some observations might be censored ($-10 \leq Diff_{i,1} \leq 10$), I estimate a tobit regression in column (2), and find that the estimate of β is still quantitatively similar.

In column (3) I estimate the model:

$$Diff_{i,t} = \alpha + \beta_1 |M_{it}| * \mathbf{1}(M_{it} > 0) + \beta_2 |M_{it}| * \mathbf{1}(M_{it} <= 0) + \varepsilon_i \tag{3.2}$$

for $t = 1$. Here $\mathbf{1}(M_{i1} > 0)$ is an indicator function that equals 1 when $M_{i1} > 0$, i.e. when $s_{i,2} > x_{i,1}$. Recall that $s_{i,2}$ is the period 1 group average, so $M_{i1} > 0$ if the subject contributed less than the group average in period 1. Therefore, β_1 captures the effect of the group choice on subjects who contributed less than the group average in round 1, and β_2 captures the effect of the group choice on subjects who contributed more than the group average in round 1. One would expect $\beta_1 > 0$ and $\beta_2 < 0$ if H_2 were true for all subjects. Columns (3)-(4) in Table C.3 show the results for this specification.

β_2 is significantly negative and β_1 is not significantly different from zero implying that subjects who contributed less than the group average in round 1 do not engage in social comparison, and continue to contribute the same low amount. Conversely, subjects who contributed more than the group average in round 1 lower their contribution by about 55 cents for every dollar that they contributed more than the social norm in the previous period. So individuals below the social norm (first period group average) stay at the same level, while individuals who were above the social norm conform to it by decreasing their contributions. This explains the lower average for round 2 reported in Table C.1.

Round 3 is similar to round 2 with the only difference being that subjects make a decision after being told that the entire group contribution distribution would be made public in the fourth round (and that identities will not be revealed). Therefore, it also corresponds to the limited information case. In the third round, 28 subjects donate at least as much as the group choice (the second period group average); 22 of them also contributed an amount equal to or more than the group choice in the second round. To check whether individual behavior changes between rounds 2 and 3, I test for the equality of the distributions of the contributions in the two rounds using the Kolmogorov-Smirnov test. The null that the two distributions are same is rejected.[12] I undertake the regression outlined in equation (3.1) for $t = 2$. The group choice in this case is the round 2 group average contribution. Columns (1)-(2) of Table C.4 show that β is still significantly positive: a one dollar difference between the individual's round 2 contribution and the group choice leads to a change of 15 cents in the individual's contribution in the

[12]However, as shown in Table 1b, I cannot reject the null hypothesis that the median difference between the two distributions is zero (the p-value of the test is 0.2184). Moreover, the K-S test is for independent samples while in my case the round 2 and round 3 contributions are clearly related.

direction of the group choice. In columns (3)-(4), I estimate the model in equation (3.2). The results are rather intriguing: $\beta_1 \approx 0.30$, while β_2 is not significantly different from zero; subjects who are above the group choice in round 2 don't change their contribution behavior, but subjects below the group choice increase their contribution, on average, by about 30 cents for every dollar that they below the social norm. This is the converse of what is observed in round 2, and also explains why the average contribution in round 3 is higher than in round 2 (Table C.1).

Round 3 is similar to round 2 in the sense that it corresponds to the limited information case. Therefore, it is not very clear whether an agent who already changed his contribution in round 2 (relative to round 1) after observing the group choice (round 1 average) will respond to the group choice in round 3 or not. I, therefore, define the dummy variable $1[\text{unchanged in round 2}]$ to equal 1 if the subject did not change his round 2 contribution (relative to his round 1 contribution), and zero otherwise. The models in equations (3.1) and (3.2) are now estimated by allowing the social comparison coefficients to be different depending on whether the subject changed his contribution in round 2 (relative to round 1). Column (1) of Table C.5 shows that there is a significant social comparison effect only for those subjects who also changed their contributions in round 2. However, column (2) shows that all subjects who were below the group choice in round 2 increase their round 3 contribution, though the increase is larger for subjects who had also changed their contributions in round 2. Conversely, subjects who contributed more than the group choice in round 2 do not seem to change their contributions in round 3 irrespective of whether they had changed their contributions in round 2 or not. These results suggest that in round 3 every subject who was below the group choice (in round 2) increases their contribution independent of whether they had changed their contribution in round 2 or not.

I interpret the results of round 2 and round 3 as follows: individuals who were below the group choice in round 2 don't seem to indulge in social comparison in round 2 since their contributions don't change relative to round 1. Some of the same individuals change their contributions in the direction of the group choice in round 3. It is hard to interpret this change as social comparison concerns since, if that were the case, one would have also observed a similar change in their contribution behavior in round 2. It seems that the prospect of the contribution distribution being made public in the next round causes them to increase their contributions, even though they know that identities will not be revealed. It could be that revelation of the contribution distribution makes their relative contribution more salient such that, for example, they might not want to be the lowest contributor in the group; this would be consistent with the theory of avoiding shame.[13] Conversely, individuals who contributed more than the group choice in round 1 engage in social comparison and decrease their contributions in round 2, while in round 3 no effect is found for such individuals since they had already adjusted their contribution in the previous round. The evidence from the analysis of rounds 2 and 3 suggests that claim 2 (i.e., contributions should move closer to the group choice because of learning about the social norm) only holds for individuals who contribute more than the group choice in the first round. It should be pointed out that contributions could change in rounds 3 and 2 (relative to the previous round) for an altruistic subject, especially if the revealed group choice causes him to update his beliefs about the total amount the charity receives

[13]See Bowles and Gintis (2003) for a theoretical model where shame increases the level of cooperation in a group.

from his group. Moreover, the change in the contribution of such an agent will depend on both the functional form of his utility function and his beliefs. Unfortunately, I cannot test this because empirically I cannot distinguish between the various types of agents. Therefore, even though there is limited support for hypothesis 2 in the data, the results should be interpreted with caution since they could partially be driven by an altruistic agent changing his contribution (in the absence of social comparison).

I next check if there is evidence for claim 3 (i.e., contributions increase in the group choice) in the data. This leads to the hypothesis:

H_3 : *Contributions are increasing in the group choice.*

I regress the contribution on the group choice which is presented to the individual when he makes his contribution decision. As can be seen in Table C.6, this hypothesis finds strong support in the data.

I now move to the analysis of the fourth round, the first round in which subjects make a decision under the knowledge that their identities will be revealed. Table C.1 shows that the highest number of subjects donate in this round, and that the contribution average is the highest in this round. Moreover, there is a positive treatment effect on round 4 contributions (Table C.2). I first check if claim 4 holds. As discussed in section 3.3.3, this requires testing the hypothesis:

H_4 : *The round 4 contribution distribution first-order stochastically dominates the round 3 contribution distribution.*

I test H_4 graphically. Figure C.1 presents the cumulative distributions of the contributions of rounds 3 and 4. As can be seen, it is indeed the case that Pr(contribution$\leq x$) for $x \in [0, 10]$ is weakly lower in round 4 than in round 3. Thus, claim 4 holds in the data. In other words, the prospect of identification of subjects causes them to contribute at least as much as they had in the previous rounds. Since the subjects observe the entire round 3 group contribution distribution, it is not clear what information they use to decide their contribution in round 4. To explain subjects' behavior in round 4, I undertake the following regression

$$
\begin{aligned}
(3.3) \qquad Diff_{i,t} &\equiv x_{i,t+1} - x_{i,t} \\
&= \alpha + \gamma_1[\text{Mode}(X_{i,t}) - x_{i,t}] + \gamma_2[\text{Mean}(X_{-i,t}) - x_{i,t}] \\
&\quad + \gamma_3[\text{Median}(X_{i,t}) - x_{i,t}] + \gamma_4 MC_i + \varepsilon_i
\end{aligned}
$$

for $t = 3$; here Mode$(X_{i,t})$ is the mode of the contributions in round t in i's group, and Mean$(X_{-i,t})$ is the mean of the contributions of other group members in period t. In this specification, for example, γ_2 captures the change in the individual's contribution from round t to round $t+1$ in response to the distance between his contribution in round t and the mean behavior of others in round t. A positive γ_2 would imply change in the contribution in the direction of the mean. MC_i is i's score on the Marlow-Crowne test; since the score is increasing in one's social desirability, one would expect $\gamma_4 > 0$.

The tobit regression result of equation (3.3) for Round 4 is presented in the first column in Table C.7. In round 4, the subject changes his contribution in the direction of the mode ($\gamma_1 > 0$), and the mean ($\gamma_2 > 0$), but away from the median ($\gamma_3 < 0$). Individuals change their round 4 contribution by 25 cents for both a \$1 gap between their round 3 contribution and the group mode, and for a \$1 gap between the round 3 contribution and the group average. The positive γ_1 is intuitive since it is logical to see conformism as an inclination towards the most frequent choice by the group members. Rather surprisingly $\gamma_3 < 0$ implying that individuals move away from the median contribution of the group. This result can be explained as follows: the mode contribution in most groups is zero, while the mean contribution is less than the median contribution for the majority of groups. In the case where individuals are inclined to conform to the modal and mean choice of the group, one would expect γ_3 to be negative. Moreover, γ_4 is statistically insignificant, which suggests that behavior is not correlated with one's performance on the M-C test.

Round 5 is similar to round 4, except that subjects are informed that their precise action (whether they donate or not) will also be revealed. One would expect the number of people who donate to increase in this round. However, Table C.1 shows that this is not the case. The average amount contributed also goes down in this round. Since a high contribution norm never evolves in round 4, subjects who only donate in round 4 may think that a contribution of zero is socially acceptable, and hence revert back to not contributing (evidence of this is documented in section C.3). Moreover, this could also be due to subjects who donated in earlier rounds expressing frustration at the lack of contribution by others, and hence decreasing their own contributions in the final round. The second and third columns of Table C.7 show the results of equation (3.3) for rounds 5 and 6 respectively. The results for Round 5 are similar to those for round 4. However, no coefficient is found to be significant in round 6. Recall that round 6 is analogous to round 5. The average amount contributed in round 6 is lower than in round 5. This finding seems to be similar to the stylized fact of repeated public goods games that contributions decline over the period of repetition (Ledyard, 1995). Column 4 of Table C.7 presents the tobit regression for the pooled sample, i.e. for rounds 4, 5, and 6 combined. I allow the error term, ε_i, to be correlated across the rounds for the same individual. The results are qualitatively similar to those obtained for the fourth round.

3.5.2. Who Affects Whom?

It might be useful to further look into the behavior of subjects in rounds where a contribution decision is made after the identities and actions of group members have been revealed. Each subject was asked to report members of their group they knew from outside the laboratory.[14] It would be interesting to see how, in a setting where strategic complementarities are absent, subjects respond to the behavior of people they know in the group, and to that of individuals who are strangers to them. As Ouchi (1981) says about behavior in groups:

> *"What we care about most is what our peers think about us... More than hierarchical control, pay, or promotion, it is our group memberships that influences our behavior. There are daily examples of the tremendous power group memberships can exert upon people to the extent of changing their religious beliefs, their attitudes towards work, and even their self-image... It is not external evaluation or rewards that matter in such a setting (the workplace), it is the intimate, subtle and complex evaluation by one's peers – people who cannot be fooled – which is paramount." (Ouchi 1981:pg. 25)*

Bandiera et al. (2007) find that presence of friends affects worker's performance in a setting where there are no externalities across workers due to the compensation scheme.

[14] This was asked in round 5 once the subject observed the names and location of each group member. 37 subjects reported that they knew at least one other person in their group from outside the lab.

Surprisingly, very few experimental studies have attempted to study the economic significance of social ties (defined as subjects knowing each other from outside the lab). For example, Haan et al. (2006) find that friends are likely to contribute more to the public good than other classmates.[15] However, I am not aware of any experimental attempt to study the influence of friends in cases when there are no externalities arising from the monetary incentive scheme across individuals. For $t = 4$ and $t = 5$, I estimate the following equation in order to see how friends affect one's choice:

$$x_{i,t+1} - x_{i,t} = \alpha + \eta_F[\overline{X_{i,t}^{Friends}} - x_{i,t}] + \eta_S[\overline{X_{i,t}^{Strangers}} - x_{i,t}] + \varepsilon_i \tag{3.4}$$

where $\overline{X_{i,t}^{Friends}}$ is the average contribution in round t of members in i's group with whom the subject has some sort of acquaintance from outside the lab, while $\overline{X_{i,t}^{Strangers}}$ is the average contribution of group members whom i does not know from outside the laboratory. η_F and η_S are the parameters of interest. One would expect $\eta_F > 0$ ($\eta_S > 0$) if the subject is concerned about being close to the choice of the friends (strangers) in the group. The results of a tobit regression of equation (3.4) are presented for decisions taken in rounds 5 and rounds 6. The results for round 5 ($t = 4$ in equation 3.4) are shown in column (1) of Table C.8: η_F is significant and positive: a one dollar gap between the average of one's friends and one's own contribution causes the subject to change his next period contribution by 10 cents in the direction of the average of the friends. No corresponding effect is found for strangers. In round 6, there is only weak evidence of contributions

[15]Other exceptions are Abbink et. al (2006), and Reuben et al. (2007). The former studies the role of social ties in an experimental group project meant to mimic the setup of microfinance institution. They find that self-selected groups exhibit a higher willingness to contribute in the beginning of the experiment. The latter explores the effect of the presence of social ties on emotional reactions in a three-player power-to-take game.

being correlated with those of one's friends. On the whole, it seems that the change in one's behavior is indeed correlated with the choices of people they know in their group.

To dig deeper into the question of how friends are affecting each other, I define $Gap^F_{i,t} \equiv \overline{X^{Friends}_{i,t}} - x_{i,t}$. The indicator $\mathbf{1}[Gap^F_{i,t} > 0] = 1$ if the subject donated less than the average contribution of his friends in round t. I estimate the following equation:

$$
\begin{aligned}
(3.5) \qquad & x_{i,t+1} - x_{i,t} \\
= \; & \alpha + \eta_{F1} * |Gap^F_{i,t}| * \mathbf{1}[Gap^F_{i,t} > 0] + \eta_{F2} * |Gap^F_{i,t}| * \mathbf{1}[Gap^F_{i,t} < 0] \\
& + \eta_S[\overline{X^{Strangers}_{i,t}} - x_{i,t}] + \varepsilon_i
\end{aligned}
$$

where η_{F1} (η_{F2}) captures the effect on i's contribution in period $t+1$ if his period t contribution was less (greater) than the period t average of the group members whom he knows from outside the lab. $\eta_{F1} > 0$ ($\eta_{F2} < 0$) would imply that an individual increases (decreases) his period $t+1$ contribution if he was below (above) the average contribution of his friends in round t. The results of equation (3.5) for $t = 4$ and $t = 5$ are shown in the columns (3) and (4) of Table C.8 respectively. In both rounds, subjects don't seem to respond to the behavior of strangers. In round 5, subjects who contributed more than their friends (in round 4) decrease their contribution ($\eta_{F2} \approx -0.15$; for every dollar that they contributed more than the average of their friends in round 4, they decrease their round 5 contribution by about 15 cents). Contributions of subjects who contribute less than their friends stay unchanged (η_{F1} is not significantly different from 0). This evidence suggests that subjects evaluate themselves relative to their friends.[16] However, it is not clear why

[16]For example, one subject wrote: "Even though group members knew my identity, I wasn't especially influenced because I did not know them." See section C.3 for other comments.

the high contributors change their contributions but not the low contributors. In round 6, $\eta_{F2} < 0$; individuals who contribute one dollar more than the average of their friends decrease their contributions by about 12 cents. To summarize: in round 5, subjects above the friend's average decrease their contribution, while those below the average keep their contributions the same. In round 6, when this pattern is made public, subjects realize that their friends are not increasing their donations, so friends decrease their contributions even more. On the whole, it seems that individuals feed off the behavior of their friends: the presence of friends in one's group generates a contagion that cause subjects to donate less in the presence of friends.

3.5.3. Discussion of Results

The purpose of this experimental setup was to disentangle some of the mechanisms through which conformity arises. Section 3.2 outlined four possible channels of conformity: (1) social learning, (2) social comparison, (3) strategic complementarities, and (4) social influence. By making the payoff of the individual dependent only on his own action, I rule out the third explanation in the setting considered in this paper. The goal is then to disentangle the other three mechanisms. The experimental setup does not allow me to distinguish between social learning and social comparison. However, it successfully distinguishes social influence from the other mechanisms.

The analysis in section 3.5.1 shows that the experimental setup yields results which confirm most of the predictions of the theoretical model outlined in section 3.3. In round 1, in the absence of any group information, the behavior of individuals is found to be correlated with their beliefs of other group members' actions. This would be the case if an individual believes that the actions of others is the *appropriate* behavior in this context, or if agents are conditional contributors. Round 2 in the experimental setup presents the most clean evidence for the case of the limited information case (subjects observe some statistic of other group members' contributions). The results indicate that conformity does in fact arise because of concerns for social comparison; however, only the subjects who contribute more than the group average in round 1 conform to the norm of low contributions by lowering their own contributions. It is not clear why the converse doesn't happen, i.e. why people below the norm do not increase their contributions. Round 3 is similar to the limited information case. However, the converse of round 2 is now observed: subjects below the norm increase their contributions and conform to a norm of a higher donation now. However, the results of this round should be interpreted with caution since individuals might also be responding to the fact that their contributions (though not identities) would be made public in the next round. An individual wanting to avoid the guilt and shame of being the lowest contributor in the group may change his offer in the third round.[17] Another possible explanation for the change in contributions in rounds 2 and 3 (relative to the previous round), which cannot be ruled out, is that it is partially being driven by altruistic agents who update their beliefs about the total amount the Red Cross receives from their group. This would change the interpretation of the results for rounds 2 and 3– conformity in that case would be a consequence of social learning (in addition to social comparison). As mentioned earlier, since the type of an agent cannot be inferred from his action, I cannot empirically test for this explanation.

[17]See Charness and Dufwenberg (2006) for an experimental investigation of how guilt aversion may be relevant for understanding interactions.

Therefore, I conclude that rounds 2 and 3 present evidence of conformity because of social comparison and (possibly) because of social learning.

Round 4 presents the cleanest evidence of the full information case (subjects make a decision knowing that both their contribution and identity will be made public in the next round). Changes in contributions in this round are attributable to image-related concerns (*social influence*). As predicted by the model, the round 4 contribution distribution first order stochastically dominates the contribution distribution from the third round. Moreover, the results indicate that the contributions of the individuals conform to the modal choice in the group. Since the decision in round 4 is made prior to the revelation of the group members, and all members face the same institutions, the conformity in contributions in this round can be interpreted as an endogenous interaction in the terminology of Manski (1993). Rounds 5 and 6 are slight modifications of round 4. In round 4-6, the analysis reveals that subjects conform to the mode: for every dollar that the subject's contribution is away from the mode, he changes his contribution by $0.25 in the direction of the mode.

Since an individual is aware of the exact composition of his group in rounds 5 and 6, the interaction effect may now be capturing both the exogenous effect as well as the endogenous effect. Subjects were asked to report which group members they knew from outside the laboratory. In section 3.5.2, I use this information to further understand the group behavior. It emerges that individuals only respond to the contributions of group members whom they know from outside the laboratory. More specifically, they move their contribution in the direction of that of their friends. However, the decrease in the contributions of individuals who contributed more than their friends in the previous round is significantly larger than the corresponding increase in the contributions of individuals who donated less than their friends in the previous round. The consequence of this is that a low contribution norm emerges that causes friends to contribute less in the presence of friends. To my knowledge, Bandiera et al. (2007) is the only other study that looks at social interactions in a setting where there are no externalities arising from monetary incentives. They find that friends conform to a common productivity norm that lies between the productivities of the most and least able friends. In the current context, one possible explanation for why the presence of friends does not provide positive role models or generate incentives to contribute to be the most generous in the group could be that subjects treat the experiment as an artificial setting, and could rationalize their actions as not being reflective of their everyday behavior; in that case, a low contribution norm rather than a high contribution norm would be more likely to be observed amongst friends.

3.6. Conclusion

The attempt of this study was to go beyond what most economic studies of social interactions do, and instead of just measuring the extent of social interactions, pin down potential channels through which conformity may arise. This paper introduces a simple model in which conformity arises from social comparison concerns and from image-related concerns (social influence). In order to disentangle the two, I use the fact that social influence only matters if actions are observable to others. The model predictions are tested in an experimental setting which unmasks individuals' actions and identity in a controlled and systematic way.

The empirical methodology developed in this paper disentangles learning about the norm (social comparison and learning) and social influence as possible channels of conformity. This distinction is relevant from both a theoretical point of view to understand the processes through which individual choices are affected, and from a policy point of view. The empirical setting considered in this paper is an experimental charity contribution game in which each individual is randomly assigned to a group. A key feature of the setup is that one's payoff only depends on their own actions, and therefore, competing hypotheses like reciprocity can be ruled out in explaining the motivations of the agents. I find that individuals indulge in social comparison and change their actions in the direction of the social norm even when their identities stay hidden. Once identities and contribution distributions of group members are revealed, individuals conform to the modal choice of the group.

The second set of findings sheds some light on how social ties affect choices of individuals. Using information provided by the individuals on which group members they know from outside the laboratory, I find that individuals only respond to the contributions of their friends. Moreover, the analysis reveals that a low contribution norm evolves that causes individuals to contribute less in the presence of friends.

The experimental setting in this study allows me get around the difficult identification problems in measuring social interaction effects in real-world instances (see Manski, 1993; 2000), and presents clean evidence on some of the mechanisms through which conformity arises. However, the design used in this study fails to disentangle social learning and social comparison. Another limitation of this study is the fact that the evidence is based on an experimental setting. Since the laboratory setting is an artificial setting where stakes are much lower, one should be careful in extrapolating the findings in this study to other settings. However, I believe that given the nature of social influence, the effects found in this study offer a lower bound for effects found in real-world instances. In the case of real-world applications, one would need richer and more specific data than is typically available in order to disentangle the various causes of conformity. I intend to explore this further in future work.

CHAPTER 4

College Major Choice: Revisions to Expectations, Perceptions of Discrimination, and Experimentation

4.1. Introduction

Individuals choose a college major under uncertainty- uncertainty about personal tastes, individual abilities, and realizations of choice-specific outcomes. Understanding any decision under uncertainty requires one to study how subjective expectations and preferences are used to make the choice. In Zafar (2007), I estimate the decision rule of college major choice by eliciting individuals' subjective expectations about major-specific outcomes. However, for credible policy recommendations, it is crucial to understand the process of expectations formation. This is the focus of the current study: I re-interview a sample of students who took the initial survey to understand how individuals process information to form expectations and revise them.

Individuals may revise their expectations about various major-specific outcomes as new information becomes available. A study of expectations revision requires longitudinal data on subjective expectations. However, understanding the mechanisms that lead individuals to revise their beliefs also requires data that directly identifies new information. This can be a challenging task since individuals may have access to various sources of information. The follow-up survey data in conjunction with the data in Zafar (2007) allows me to conclude that changes in expectations about various major-specific outcomes vary in sensible ways. This matches with conclusions reached in Bernheim (1990), Dominitz (1998), Hurd and McGarry (2002), Delavande (2007), and Lochner (2007) who find that expectations are responsive to new information.[1] However, existing studies, due to lack of data that identifies new information, cannot pin down the causal explanation for the revision in expectations.[2] In this paper, data on GPA beliefs at different points in the future and their realizations allows me to assess the responsiveness of GPA expectations to new information. Individuals who receive positive information increase their prediction of short-term future GPA, while individuals who receive negative information only revise their predictions downward if the information content is very negative. Moreover, I don't find any effect on long-term GPA expectations.

The results in this paper also contribute to the body of positive evidence that has been accumulated on the validity of subjective data (see Manski, 2004, for an overview of the literature). I find that priors for outcomes like approval of parents, and graduating with a GPA of more than 3.5 are fairly precise, and that individuals don't revise them by as much as they revise their priors for outcomes that are realized in the workplace. These findings are consistent with students adopting a Bayesian learning approach; for outcomes associated with college, one would expect students to have fairly precise information at the time of the initial survey. Conversely, for outcomes in the workplace, one would expect students to receive useful information between the two surveys. The panel on

[1]Delavande's (2007) contribution is more of a methodological nature; she develops a method to measure revisions to subjective expectations about binary outcomes. Bernheim (1990), Dominitz (1998), Hurd and McGarry (2002), and Lochner (2007) study revisions to expectations about social security benefits, income, survival, and arrest respectively. In this study, I focus on revisions to expectations about several major-specific outcomes.

[2]Some laboratory experiments have studied how agents update their beliefs with new information; see, for example, El-Gamal and Grether (1995), and Houser, Keane and McCabe (2004). However, they use extremely stylized settings, and study how agents learn over short time horizons.

subjective beliefs also allows me to answer certain doubts raised about the validity of subjective expectations data (Bertrand and Mullainathan, 2001). For example, in my data, I can rule out cognitive dissonance as a potential issue. Cognitive dissonance would imply that subjects report attitudes that are consistent with their behavior (for example, if they never pursue a major, they tell themselves that they never liked it anyway). So, one would expect larger unfavorable changes in beliefs for outcomes in majors that an individual never pursued, and similarly larger favorable changes in beliefs for outcomes in the major that the individual has stuck with. However, in this study, I don't find any empirical support for this.

This study also focuses on explaining the reasons for the gender gap in fields like Engineering. More specifically, in light of the findings in Zafar (2007), I try to understand why females are less likely to enjoy studying fields like Engineering. For this purpose, I elicit beliefs of both monetary and non-monetary discrimination associated with the various majors. Both males and females seem to be aware of a positive wage gap in favor of males in most fields. However, they tend to underestimate the extent of the wage gap, and *incorrectly* believe that the wage gap stays roughly constant over time (realizations indicate that the gap increases over time). Moreover, more males than females attribute the wage gap to "characteristics and aptitudes actually being different between males and females", while a larger fraction of females state "employers expecting different characteristics between males and females" as one of the main reasons for the wage gap. Such a combination of beliefs, as shown in Filippin (2003), can be self-confirming in a game-theoretical equilibrium and can lead females to underinvest.

The survey also elicited individuals' beliefs about males and females being treated poorly in the jobs in various majors. Perceptions of being treated poorly in the jobs in a given major are found to be negatively correlated with the fraction of people of the same gender majoring in the field, with the gender gap in wages in the field, with beliefs of enjoying coursework, and with enjoying working at the jobs. The analysis reveals that the largest difference in females being treated poorly relative to males is in Engineering and Math & Computer Sciences (two categories with the lowest fraction of females). Social psychology studies have shown that occupational segregation by gender can be explained by the emphasis of certain gender-specific attributes in a given occupation (Anker, 1997). Cejka and Eagly (1999) find a positive correlation between female-dominated occupations and the perception that feminine attributes are essential for success in those fields. However, it is not clear how to interpret the positive correlation between beliefs of females being treated poorly in the jobs and fraction of females taking classes in that major; it could be that females prefer fields that value female-specific attributes and where females are treated more favorably, or it could be that females are treated more favorably at those jobs precisely because those are "female" occupations. I re-estimate the choice model (initially estimated in Zafar, 2007) by including this variable, but the estimation results qualitatively remain the same since the variable is positively correlated with beliefs of enjoying coursework and enjoying working at the jobs.

Finally, this paper tries to understand why individuals may experiment with different majors. There is a theoretical literature which emphasizes that individuals may find it optimal to experiment with different majors to learn about their ability and match quality (Manski, 1989; Altonji, 1993; Malamud, 2006). However, this aspect has not been investigated empirically.[3] With access to limited data, I conduct a descriptive analysis of students' experimentation with majors. Nearly 60% of my sample reports to have experimented with at least one other major. Academic performance does not seem to be the only reason for major switches. Losing interest in the original major, and getting interested in something else are reported to be the main reasons for switching majors.

The paper is organized as follows: Section 4.2 outlines the data collection methodology, and describes some of the subjective data in detail. Section 4.3 presents the perceptions of monetary and non-monetary discrimination. Section 4.4 focuses on revisions to expectations. Section 4.5 undertakes a descriptive analysis of experimentation with majors. Finally, Section 4.6 concludes.

4.2. Data

156 of the 161 respondents who had taken the initial survey (discussed in Zafar, 2007) had agreed to be contacted for the follow-up. Individuals who had given their consent were contacted by E-mail for the follow-up; the E-mail summarized the findings of the initial survey and the purpose of the follow-up. Students were told that they would be compensated $15 for the 1-hour electronic survey. The follow-up was conducted between November 2007 and February 2008 in the PC Laboratory located in the Northwestern Main Library.

Of the 156 initial survey respondents, 117 (75%) took the follow-up survey. The first column of Table D.1 shows the characteristics of individuals who took the follow-up survey. For comparison, characteristics of the initial sample and the actual sophomore

[3] An exception is Malamud (2006). In his model, learning about match quality in different fields of study provides the individual with information on match quality in the occupations related to those fields. His empirical investigation focuses on the effect of early versus late specialization.

population are shown in columns (2) and (3) respectively. Follow-up survey respondents seem to be similar to the initial survey respondents in most aspects. Even though the average GPA of follow-up respondents is higher than that of the initial survey-takers, the difference is not statistically significant. Students of Asian ethnicity are over-represented in the sample relative to their population proportion. Survey-takers, especially males, have higher average GPAs than the sophomore population.

The survey consisted of four parts. The first part collected both qualitative as well as quantitative data on experimentation with majors. The second part elicited beliefs for major-specific outcomes for three different major categories in the individual's choice set; beliefs about the major-specific outcomes were elicited for: (1) the major that the individual was actually pursuing, (2) the individual's second major (or the second most preferred major if the student did not have a second major), and (3) a major that the individual had once pursued but was no longer pursuing (if this was not applicable, beliefs were elicited for the least preferred major in the individual's choice set). The purpose of the second part of the survey was to study the evolution of beliefs. The third part collected perceptions of monetary and non-monetary discrimination in the various majors. The last part of the survey collected data on the individual's GPA at different points in the past, as well as their beliefs about their academic performance at different points in the future. Individuals were also requested to upload their transcript; only 41 (35%) respondents gave access to their transcript data. At the end of the survey, respondents were asked if they were willing to be contacted for a potential follow-up in the future- 112 (96%) survey-takers gave their consent.

4.2.1. Subjective Beliefs About Starting Salaries

Survey respondents were asked about the average annual starting salary of Northwestern Bachelor Graduates of 2007 for up to three different major categories. As outlined in Zafar (2007), the purpose of this question was twofold: (1) the responses can be directly compared to actual salary realizations of Northwestern graduates, and (2) since large earnings premiums exist across majors (Arcidiacono, 2004), I can test whether respondents are aware of these income differences. The problem with the initial survey was that the question was vague- it wasn't clear whether the point estimate that respondents provided was a point on their gender-specific subjective earnings distribution, or a point on the general earnings distribution. This problem was rectified in the follow-up survey. For three different major categories, the respondent was asked the average starting salary for both genders. The question was: *"What do you think was the average annual starting salary of Northwestern G graduates (of 2007) with Bachelor's Degrees in X?"* where $G = \{$Male, Female$\}$. Though there is substantial heterogeneity in the beliefs, I only present the mean responses in Table D.2. Analysis of the first six columns of Table D.2 shows that respondents are aware of different returns to majors; the relative subjective beliefs seem to be consistent with actual trends. There are, however, a few notable patterns. Both males and females underestimate the average salaries (for both genders) for all categories except Natural Science and Ethics and Values. However, compared to their male counterparts, female respondents report higher average starting salaries (for both themselves as well as for males) for Engineering and several WCAS majors. Column (7) shows that the realized wage gap is in favor of males for all WCAS categories except Area Studies and Literature & Fine Arts. The survey respondents (both males and females), on average, believe that the wage gap is in favor of males for all major categories (columns 8 and 9). However, both males and females tend to underestimate the extent of the gender gap in wages for most majors.

Table D.2 only shows the average beliefs by gender. Using the demographic information collected from the respondents, I can possibly say something about the determinants of the errors in the respondents' beliefs about Northwestern 2007 graduates' salaries. As in Betts (1996) and Zafar (2007), I use the following metric to model the respondents' errors:

$$\ln\left|\frac{\overline{s_{im}^{G}} - s_{obs\_m}^{G}}{s_{obs\_m}^{G}}\right|$$

where $\overline{s_{im}^{G}}$ is respondent i's reported average starting salary in major m for gender G ($G = \{\text{Male, Female}\}$), and $s_{obs\_m}^{G}$ is the true average salary for Northwestern 2007 graduates of gender G in major m. Table D.3 shows the results of regressing this metric on various demographic characteristics. A random effect is included for each respondent in order to account for random differences in estimates between the respondents. The first column reproduces the results from the first survey; however, since the follow-up survey only elicits beliefs for three different majors in the individual's choice set, I restrict the analysis in the first column to the same three majors. In order to understand whether the genders differ in any systematic ways in which they make errors, I also present the results separately for cases where the respondents' point estimates are greater (less) than the observed outcomes. Students with higher GPAs appear to make significantly smaller errors when estimating starting salaries. On the other hand, individuals with higher SAT Math scores make larger errors, while students with higher SAT Verbal scores make smaller errors. One would expect that individuals majoring in a given field would have better information about their chosen field. The regression includes a variable "Studying Major" that equals 1 if the student is studying the major about which she reports the starting salary. However, the coefficient on this variable is insignificant. Individuals who were studying the given field for which they reported the starting salary and had also declared the major at the time of the initial survey make smaller errors (though the coefficient is statistically insignificant). This would be consistent with a story where information acquisition is costly, and individuals only seek information about a major when they are fairly sure about pursuing it. Individuals with a college-educated father make significantly smaller errors; this is consistent with students with college-educated parents having access to more precise information. I don't find evidence of individuals with parents who have studied a given major being better-informed about starting salaries in that major. One of the more notable findings is that females make significantly larger errors; on average, they make errors that are about 40% larger than those of males. This finding contrasts with Betts (1996) who does not find any statistically significant difference in the error patterns between males and females. Female respondents in the current study make large errors in the starting salaries for both males and females, and are more likely to overestimate them. Therefore, it doesn't seem that they are only self-enhancing the starting salaries for their own gender. This result contrasts with Smith and Powell (1990) who do not find any statistically significant difference between the earnings expectations of college graduates between males and females.

4.2.2. Subjective Beliefs about Labor Force Participation

The follow-up survey elicited respondents' beliefs about both their full-time and part-time labor force participation at the ages of 30 and 40. There is substantial heterogeneity in the beliefs between males and females, as well as within each gender group. The lower panel of Table D.4 only shows the average labor force beliefs of males and females. The female mean belief of full-time labor force participation at the age of 30 is 81.45% versus 91.75% for males. Females exhibit a greater heterogeneity in their beliefs (a standard deviation of 19.10 versus 8.80 for males). Conversely, females have significantly higher beliefs (as well as greater heterogeneity) about part-time labor force participation and being unemployed at the age of 30. Labor force participation beliefs were also elicited for the age of 40. Similar trends emerge in those beliefs. One notable feature is that labor force beliefs at the age of 40 are similar to those at the age of 30. The U.S. Bureau of Labor Statistics projects the participation rates of males and females of ages 30-34 in 2016 to be 96.6% and 74.7% respectively. The corresponding numbers for ages 40-44 is 90.7% and 75.6%. Even though Northwestern undergraduates are a very selective group, their responses do compare favorably with the predicted labor force beliefs.

I next see whether the heterogeneity in full-time labor force beliefs is associated with the demographic characteristics of the respondents. The first 3 columns of Table D.5 (Table D.6) presents best linear predictors under square loss of the labor force participation rates at the age of 30 (40). The belief of being active in the full-time labor force for females is, on average, about 12 points lower than that of males at both ages. Another notable finding is that the labor force participation beliefs at the age of 30 (40) of females with a stay-at-home mother are about 5 (17) points lower than those of females with a working mother; no corresponding effect is found for males. This evidence is consistent with the intergenerational learning model about payoffs to work for females developed in Fernandez (2007); in her framework, females receive private and public signals through which they learn about the payoffs to work. Here, it seems that females give a lot of weight to the private signals they receive from their mothers. Fogli and Veldkamp (2007) also develop a similar model where female labor force participation increases through learning from endogenous information.

The top panel of Table D.4 reports the *average* fertility beliefs at the age of 30 and 40 for males and females. The two genders have similar average beliefs; both expect, on average, to have one child at the age of 30, and 1.75 children at the age of 40. Survey respondents were also asked about their belief of being the primary bread-earner of their family (defined as contributing the larger proportion of total household income) at the age of 30 and 40. Average responses are shown in Table D.4. At both ages, the average belief of being the primary bread earner for males is about 75%, and about 50% for females (the gender difference is significant at 0.01). The table also reports the average beliefs about the fraction of time the respondent expects to spend on house work at the age of 30 and 40; females, on average, expect to spend at least 30% more time on house work relative to men. It is not very clear how to interpret these beliefs. For example, females could have lower beliefs of being the primary bread-earner in the family either if they don't plan to be active in the labor force, or if they believe that their partners would be earning more than them and, hence, they are more likely to devote their time to home production.[4] Nonetheless, I use these additional variables to see how full-time labor force participation beliefs vary with various individuals characteristics. The last 3 columns of Table D.5 (Table D.6) presents best linear predictors under square loss of the labor force participation rates at the age of 30 (40) with the inclusion of these new variables. At the age of 30, the coefficient on the fraction of time spent on home production is large and significant. The coefficient on the female dummy decreases in magnitude by about 5 points. Beliefs of being the primary bread-earner, and time spent on home production both significantly affect labor force beliefs at the age of 40. The coefficient on the female dummy is now insignificant. However, the effect of a stay-at-home mother continues to be large and significant (which indicates that the channel through which they affect females' beliefs about labor force participation is not captured in any of the newly added variables). Expected number of children does not seem to affect one's labor force beliefs significantly. It is hard to interpret the coefficients on these variables since the causality can go in either direction.

4.3. Beliefs of Discrimination

One of the main findings in Zafar (2007) was that the gender gap in majoring in fields like Engineering was not because of gender differences in beliefs about academic ability or future expected income. Conversely, most of the gap was because of gender differences in beliefs about enjoying studying the various majors and gender differences in preferences. However, it was not clear why females believe they would enjoy studying fields like Engineering less than males. One reason for this could be innate gender differences in attitude

[4] The second explanation would be consistent with a model where females consider education as a form of pre-martial investment which enhances their chances in the marriage market (see references in Lafortune, 2008).

(Baron-Cohen, 2003). A second explanation is (monetary and non-monetary) discrimination (Valian, 1998). A third explanation is the *role model hypothesis*. The argument is that females may avoid male-dominated fields due to gender-based discrimination, and female faculty may mitigate this effect. Canes and Rosen (1995) and Bettinger and Long (2005) analyze the effect of teacher gender on student choices in college. The former study finds no evidence of female faculty affecting choices of women, while the latter finds some support for the role model hypothesis. Another possible explanation is that females enjoy studying a field more when it has more female students. It is hard to find empirical evidence of this since, by construction, fields that have fewer females are the ones that females tend to enjoy less. I focus on an empirical investigation of the second explanation i.e. gender-based discrimination; the follow-up survey contained several questions that elicited beliefs of both monetary and non-monetary discrimination.

4.3.1. Beliefs of Monetary Discrimination

The follow-up survey elicited the *explicit* gender gap in earnings in each major category at two different points in time. The explicit gender gap at the age of 30 was computed from the respondents' answers to the following two questions:

> (1) What do you think is the average amount of money an individual with the same characteristics and gender as you COULD earn ANNUALLY at the AGE OF 30 if they graduated with a major in X?
>
> (2) What do you think is the average amount of money an individual with the same characteristics as you BUT of the opposite gender COULD earn ANNUALLY at the AGE OF 30 if they graduated with a major in X?

The first two columns of Table D.7 show the explicit gender gap at the age of 30 as reported by males and females respectively. Ideally, one would compare these numbers with realizations of previous Northwestern cohorts at the age of 30. Unfortunately, such a comparison is not possible since Northwestern does not follow its alumni. For illustrative purposes, this explicit gender gap can be compared to the wage gap *implicit* in the starting salaries of 2007 Northwestern graduates (column 7 of Table D.2). The explicit gender gap depicted in Table D.7 is smaller than the implicit gender gap for most major categories. One possible reason for this discrepancy could be that the explicit gender gap is computed for individuals with similar characteristics, while the implicit gender gap is computed for individuals with different characteristics. The explicit wage gap reported by females is smaller than that reported by males for most categories. Moreover, for some fields (like Social Sciences II, Natural Sciences and Math), females report a wage gap in favor of themselves which does not seem to be consistent with realizations of Northwestern 2007 graduates (or salaries one year after graduation of respondents in the 1993/2003 Baccalaureate and Beyond Survey).

In the absence of a good reference group, one could possibly check if the trends in the explicit gender gap reported by survey respondents over time match up with actual patterns. The explicit wage gap at the age of 40 is reported in columns (3) and (4) of Table D.7. One feature that stands out is that, for both male and female respondents, the average wage gap at the age of 40 is not too different from that reported at the age of 30 (this finding is similar to that of Filippin and Ichino (2005) who analyze the beliefs of gender gap in wages of Italian graduates). This contrasts with existing empirical evidence which shows that the male-female wage gap increases with the workers' time in the labor force (Laprest, 1992, and Light and Ureta, 1995). I am not aware of a study that analyzes the gender difference in wage growth across majors over time. Using the 1993/2003 Baccalaureate and Beyond Longitudinal Study (B&B:93/03), I computed the gender gap in wages for the various major categories one year after graduation and 10 years after graduation. For all major categories, except Engineering, Mathematics & Computer Studies and Education, I find that the gender gap in wages increases over time.[5] These results seem to indicate that students have a misperception about wage gaps later in the career since they expect the wage gap to stay roughly constant while realizations indicate a larger wage gap over time for most major categories.[6]

Survey respondents who reported different expected earnings for males and females with similar characteristics (i.e. their responses to questions 1 and 2 above were different) were asked to explain the reasons for that. More specifically, a list of reasons was provided in the survey and the respondent was asked to provide weights to each of the reasons such that they added up to a hundred. The list of reasons was taken from the questionnaire used by Filippin and Ichino (2005). Table D.8 reports the average weights given to each of the explanations by the two genders. Both males and females report employer discrimination to be the main reason for the wage gap (average weight of about 30). There are no gender differences in the fraction of the wage gap attributed to different distribution of household duties (both genders assign it a weight of about 20%). While males believe that about 20% of the gap is because of innate differences between males and females, females only attribute 10% of the gap to this source- this gender gap is significant (p-value=0.035).

[5] Results available from the author on request.

[6] This result should be taken with caution because the realizations data is coming from the B&B:93/03 study. Northwestern undergraduates, being a very specific demographic, may not be very similar to the B&B sample. I tried restricting the B&B sample to selective institutions but the sample is a lot smaller.

Conversely, females, relative to males, assign a significantly higher proportion of the wage gap to employers expecting males and females to have different characteristics (p-value=0.022). Therefore, it seems that females are more likely to think that the gender gap in wages is because of social attitudes and discrimination. Such a combination of beliefs, whether correct or incorrect, could lead females to underinvest in activities in the labor market.

4.3.2. Beliefs of Non-Monetary Discrimination

The initial survey did not elicit any perceptions of non-monetary discrimination from the respondents. In a quest to understand why females are less likely to enjoy studying fields like Engineering, survey respondents were asked their beliefs about each gender being treated poorly at the jobs that would be available in the different major categories. The question was worded as follows: "*What do you think is the percent chance that males (females) would be treated poorly in jobs that are available in each of the following fields?*". Before analyzing the responses to this question, in columns (5) and (6) of Table D.7, I report the fraction of females that survey respondents believe take classes in the various majors. Column (7) reports the average number of females who graduated in the various majors in 2005 and 2006 (source: IPEDS 2005 and IPEDS 2006). The responses show that survey respondents are aware of 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 (8)-(11) of Table D.7. Several notable patterns stand out. One, male respondents believe that females are treated more poorly than males in jobs in all fields except Education, Literature & Fine Arts, and Music Studies; these three fields correspond to the three most female-dominated fields (in college) as reported by males in column (5) of the table. Second, females believe that they would be treated more poorly than males at jobs in all fields except Education- the field that females believe has the highest fraction of females. Third, for both the male and female respondents, the largest difference in females being treated poorly relative to males is for Engineering and Math & Computer Sciences- two categories with the lowest fraction of females (as reported by both males and females). Finally, both males and females believe that Education is the category in which males would be treated the worst.

Table D.9 presents the correlation patterns between some of the variables shown in Table D.7 and beliefs about enjoying coursework and enjoying working at jobs (which were elicited in the initial survey) for females and males separately. There is a significant correlation of about -0.35 between the beliefs of females being treated poorly at the jobs and the fraction of females in the classes in the major for both males and females (Table D.7 showed a similar pattern). Moreover, for both male and female respondents, there is a significant positive correlation between the beliefs of females being treated poorly at the jobs and the wage gap (in favor of men) in those jobs. The fraction of females in the class has a higher positive correlation with beliefs of enjoying coursework than with enjoying work at the jobs. The female belief of enjoying coursework is strongly positively correlated with both enjoying working at the jobs and males being treated poorly in the jobs, and negatively correlated with females being treated poorly. The male belief of enjoying coursework is strongly positively correlated with enjoying working at the jobs, and negatively correlated with females being treated poorly at the jobs. The correlation between the male belief of enjoying coursework and males being treated poorly is not significantly different from zero. Finally, there is a significant negative correlation between beliefs of enjoying working at the jobs and the beliefs about your own gender being treated poorly at the jobs.

These results suggest that the perception of being treated poorly in the jobs is related to the fraction of the same gender taking classes in that field, with the wage gap in the field, and beliefs of enjoying taking courses and working at jobs available in that field.[7] The negative correlation between beliefs of females being treated poorly in the jobs and the fraction of females taking classes in that major is consistent with the occupational segregation by gender, i.e. the empirical fact that majority of men and women work in occupations that are described as "male" and "female" respectively (Anker, 1997). However, it is not clear why there is a negative correlation between the fraction of females in a given discipline and beliefs about them treated poorly: 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.

4.3.3. Re-estimating the Choice Model

I re-estimate the single-major choice model that was initially estimated in Zafar (2007). The main purpose of doing this is to see how the inclusion of the new variable *"females/*

[7]Individuals were asked to explain what "being treated poorly" meant to them. See Appendix A for selective comments.

males treated poorly at the jobs" affects the parameter estimates. The same set of assumptions is made on the specification, i.e.

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

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

For the purposes of the estimation, I use the stated preference ordering of the respondents as the dependent variable. Moreover, I use the entire sample for estimation, and do not make a distinction between individuals with a single major and those with a double major. The first three columns of Table D.10 show the parameter estimates for the case where the model does not include the new variable. The results are similar to those in Zafar (2007); the difference in utility levels is positive and largest for enjoying coursework, approval of parents, and enjoying working at the jobs. For males, the difference in utility levels is largest for enjoying coursework, approval of parents, and the social status of the jobs. Conversely, for females, the three most important determinants are enjoying coursework, enjoying working at the jobs, and approval of parents. The parameter estimates of the *extended* model (which includes the new variable "females/ males treated poorly") are shown in the last three columns of Table D.10.[8] The inclusion of the new variable does not improve the explanatory power of the model for the entire sample and for females; relative to the initial model, the Wald χ^2 (a measure of goodness-of-fit which compares the likelihood ratio chi-squared of the model to one with the null model) does not change by much. Moreover, the estimates for the other determinants that were already included in the initial model stay almost the same. For females, the difference in utility levels for

[8]The underlying assumption is still that the utility is linear and separable in the various outcomes.

females or males being treated poorly at the jobs is negative, but the coefficient is not significantly different from zero. In the case of males, the difference in utility levels for females being treated poorly at the jobs is positive but not significantly different from zero; however, the coefficient on males being treated poorly is (surprisingly) positive and significant.

The parameter estimates presented in Table D.10 are hard to interpret because of the non-linear nature of the model. In order to gain insight into the relative importance of the various outcomes in the choice, I use the same metric as outlined in equation (2.14) in Zafar (2007). To be more precise, 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:

$$
\begin{aligned}
& M_{X_1} \\
\equiv\ & \left\| \overline{\Pr(choice\ = j|\ \{\widehat{\beta_1}, \widehat{\beta_2}\}} \ - \ \overline{\Pr(choice\ = j|\ \{\widehat{\beta_1} = 0, \widehat{\beta_2}\}} \right\| \\
=\ & \sqrt{\sum_{j=1}^{8} \left[\sum_{i=1}^{N} \frac{\Pr(choice\ = j|\ \{\widehat{\beta_1}, \widehat{\beta_2}\})}{N} - \sum_{i=1}^{N} \frac{\Pr(choice\ = j|\ \{\widehat{\beta_1} = 0, \widehat{\beta_2}\})}{N} \right]^2}
\end{aligned}
$$

where the first term is the average probability of majoring in choice j predicted by the model, and the second term is the average predicted probability of majoring in j if outcome X_1 were not considered. The difference of the two terms is a measure of the importance of X_1 in the choice. The *relative* contribution of X_1 to the choice is then $R_{X_1} = \frac{M_{X_1}}{M_{X_1} + M_{X_2}}$. Table D.11 presents the results of this metric. The first three columns show the decomposition results of the model excluding the variable "males/ females treated poorly at the jobs". The results are consistent with the findings in Zafar (2007): non-pecuniary attributes explain about 90% of the choice for females, and about 55% of the choice for males. Males and females primarily differ in their preferences in the workplace, with males caring more about the pecuniary aspects of the workplace, and the females valuing the non-pecuniary aspects of the workplace more. Columns (4)-(6) show the result of the *extended* model. The new variable only explains about 3% of the choice for the aggregate sample, and for females. On the other hand, nearly 8% of the choice for males is explained by this variable. On the whole, this new variable does not add much to the model. Since it is strongly correlated with beliefs of enjoying coursework and enjoying working at the jobs (Table D.9), it seems that it's impact on the choice is already being captured indirectly.[9]

The discussion so far as focussed on the estimation of revealed preference parameters. Survey respondents were also asked to state their preferences for various determinants in their choice. More specifically, respondents were asked to assign an integer between 0 and 100 to a list of reasons such that the numbers added up to a 100. Table D.12 shows the average weights assigned to the various reasons by males and females. I interpret these numbers as the relative importance of the given reason in the choice. Enjoying working at the jobs and learning more about things that interest me were the two most important reasons for choosing a major for both males and females. However, females, on average assign higher weights to this reason (the gender difference is significant). For males, the third most important stated reason for choosing a major is getting a high-paying job. Conversely, doing well in the coursework is the third most important reason for females.

---

[9]Indeed the variable "females treated poorly at the jobs" only shows up significantly (at the 1% and 10% level respectively) for females and the entire sample in a model that excludes both enjoying coursework and enjoying working at the jobs.

These stated preferences for various outcomes are consistent with the parameter estimates discussed earlier in this section, and with the results in Zafar (2007). Determinants like fraction of people of the same gender taking classes in the major or working at the jobs don't seem to be important. One surprising finding is that males, on average, are more likely than females to have been encouraged by a mentor/ role model to choose a major (similarly females assign lower weights to peer pressure, siblings making the same choice, and parents wanting them to make the choice; the gender difference is not significant for any of these reasons though). This finding is in contrast to the literature in social psychology which finds that females are more influenceable (Eagly, 1978), and that females, in contrast to males who have mainly independent self-schemas, have interdependent ones (Cross and Madson, 1997). In light of this evidence, economic studies that empirically investigate the role model hypothesis have only focused on females (an exception is Bettinger and Long (2005) who do find evidence of role models for males in a few disciplines-education and business). However, the result in this study could be driven by the fact that it is restricted to students who have at least one major in the College of Arts and Sciences (which contains majors mostly dominated by females).

4.4. Updating Beliefs

The follow-up survey elicited the beliefs of the major-specific outcomes for the individual's actual major, for the individual's second major (or the second most preferred major if the student did not have a second major), and for a major that the individual had dropped (if this was not applicable, beliefs were elicited for the least preferred major in the individual's choice set). One of the purposes of collecting this data was to study the revisions in expectations for various major-specific outcomes. However, a panel data of subjective beliefs is also useful in answering several objections raised by skeptics of subjective data. For example, Bertrand and Mullainathan (2001) list cognitive problems, social desirability, non-attitudes (i.e. attitudes may not exist in a coherent form and, when asked a question for which they lack an attitude, people may say something meaningless), cognitive dissonance, and white noise error as reasons to be cautious of subjective data. A proper wording of the survey questionnaire can get around several cognitive problems. Response distortion due to social desirability can be mitigated by either making the questionnaire anonymous, or by making respondents answer the survey online so that they don't have to answer directly to an interviewer. With regards to the issue of white noise, there is evidence that individuals round off their responses (to the nearest 5 or 10) to percent-chance questions. However, the researcher can infer the respondent's rounding practice, and interpret the numerical responses as intervals (Manski and Molinari, 2008). Moreover, in the last decade or so, economists have successfully used subjective data at face value to explain behavior in different contexts (see Manski, 2004, for an overview of this literature). For example, Van der Klauuw (2000) shows that the use of expectations data in a dynamic model of teacher career decisions improves the precision of the parameter estimates. A panel of subjective beliefs allows one to look deeper into issues of cognitive problems (like respondents making little mental effort in answering the questions), non-attitudes (individuals being unable to make a reasonable probability assessment of the relevant question because those attitudes may not exist in a coherent form), and cognitive dissonance.

Table D.13 reports the mean change in the belief of a given outcome (disaggregated by how the individual ranks the major). The table also reports the standard errors in the mean change of beliefs, and the fraction of responses which have remain *unchanged* since the initial survey in square brackets.[10] The mean change in beliefs for almost all the binary outcomes is less than 10. Moreover, the beliefs for binary outcomes and hrs/week remain unchanged for a substantial fraction of respondents. This seems to indicate that individuals answer meaningfully and carefully, and that non-attitudes is not a serious issue.[11] Cognitive dissonance would imply that subjects report attitudes that are consistent with their behavior (for example, if they never pursue a major, they tell themselves that they never liked it anyway). So, one would expect larger unfavorable changes in beliefs for outcomes in majors that an individual never pursued, and similarly larger favorable changes in beliefs for outcomes for the major that the individual has stuck with. However, the numbers reported in Table D.13 don't show any evidence of this: average changes in beliefs of outcomes in an individual's least preferred major and current major are not too different from those in other major categories. These results, on the whole, are supportive of the use of subjective data to understand behavior of individuals. I next look into the issue of how individuals revise subjective expectations.

[10] I consider the belief of an outcome to have remain unchanged if: (1) the change in beliefs is less than 10 points (on a scale of 0-100) for binary outcomes; (2) the change in beliefs is less than 5 for hrs/week spent on coursework or job; (3) the change in beliefs for salary is less than $1000.

[11] One can also check for the presence of non-attitudes by seeing if the response of 50% is the most frequent one. According to Bruine de Bruin et al. (2000), individuals may report the response 50% when they have not made a reasonable probability assessment of the question, and want to say "50-50 chance". However, I do not find strong evidence of this in Zafar (2007), or in this study.

4.4.1. Revisions of GPA Beliefs

This section outlines a simple model of belief updating. Let X_{it} be individual i's expectation at time t about the value of a variable $\mathbf{X}$ which would be realized at some point in the future. Moreover, let Ω_{it} denote individual i's information set at time t. For simplicity, I assume that $\mathbf{X}$ is a binary event so that:

$$X_{it} = E(\mathbf{X}|\Omega_{it}) = \Pr(\mathbf{X} = 1|\Omega_{it}) \tag{4.1}$$

Similarly, X_{it+1} is i's expectation about the value of $\mathbf{X}$ at time $t+1$. Individuals are assumed to use all available information in forming expectations; therefore, revisions of expectations are solely determined by new information. I further assume that, at time $t+1$, the individual has access to all information that was available at time t. Therefore, $\Omega_{it+1} = (\Omega_{it},\, \omega_{it+1})$ where ω_{it+1} is *new* information that becomes available to i between time t and $t+1$. It follows that:

$$E(X_{it+1}|\Omega_{it}) = E[E(\mathbf{X}|\Omega_{it}, \omega_{it+1})|\Omega_{it}] = E(\mathbf{X}|\Omega_{it}) = X_{it} \tag{4.2}$$

which implies that:

$$\Pr(\mathbf{X} = 1|\Omega_{it+1}) = \Pr(\mathbf{X} = 1|\Omega_{it}) + \varepsilon_{it+1} \tag{4.3}$$

where $E(\varepsilon_{it+1}|\Omega_{it}) = 0$, i.e. ε_{it+1} is a function of new information that becomes available after time t. Equation (4.3) states that the change in expectations between time t and $t+1$ about some event $\mathbf{X}$ that is realized at some point in the future is a function of new information that becomes available after time t. In the context of this study, period

t refers to the first survey, i.e. Fall 2006 and period $t+1$ refers to the follow-up survey i.e. Fall 2007 (see Figure D.1 for a visual depiction of the timeline).[12] $\mathbf{X} = 1$ refers to the binary event that the GPA at the end of Spring 2008 (which is realized after the individual takes the follow-up survey) is above a certain threshold.[13]

So, $\Delta \Pr(\mathbf{X} = 1|\Omega_{it+1})$ (defined as $\Pr(\mathbf{X} = 1|\Omega_{it+1}) - \Pr(\mathbf{X} = 1|\Omega_{it})$) is the change in i's subjective belief about her Spring 2008 GPA being above a certain threshold between the Fall 2006 and Fall 2007 surveys. Panel A of Figure D.2 depicts the local polynomial (of order 1) estimates of the regression of change in the Spring 2008 GPA beliefs on the change in the individual's GPA between the two surveys.[14] Revisions of Spring 2008 GPA expectations seem to be positively related to changes in realized GPA. The change in beliefs about Spring 2008 GPA in response to positive and negative changes in realized GPA is almost symmetric. Similar responsiveness to positive and negative changes in realized GPA may lead one to conclude that increases and decreases in realized GPA between the two surveys contained equally useful information. However, to be able to conclude this, one needs to discern the information content of the GPA realized at the beginning of Fall 2007. More specifically, one needs to know the respondent's prior probability distribution (i.e. their belief in Fall 2006) about their GPA at the start of Fall 2007.[15] In the absence of this information, one may conclude positive information for negative information

[12]The initial survey spanned the period from November 2006 to February 2007, but I will denote it as Fall 2006 in the empirical analysis. Similarly the follow-up survey spanned the period from November 2007 to January 2008, but I will denote it as Fall 2007.

[13]In this case the threshold is the individual's GPA at the beginning of Fall 2006.

[14]I use a local linear regression estimator instead of a Kernel regression since this avoids the boundary problem. I experimented with different bandwidths but the figures did not change much.

[15]To be more precise, the change in GPA between the two surveys actually is the difference in GPA at the beginning of Fall 2007 (which would be the GPA realized at the end of Spring 2007) and the GPA at the beginning of the quarter when the individual took the initial survey. Therefore, realized Fall 2007 GPA actually means the GPA realized at the end of Spring 2007.

when the individual's GPA in Fall 2007 decreases but by less than the individual had anticipated.

Panel B of Figure D.2 depicts the local linear regression estimates of the change in graduating GPA on changes in realized GPA between the two surveys. Both surveys elicited the individual's belief of their GPA at the time of graduation in their major being above a certain threshold; the dependent variable is now the change in this belief. As depicted in panel B, individuals do not increase their belief of graduating GPA in response to positive changes in realized GPA. In order to understand the responsiveness of beliefs in GPA at different points in the future, I try to discern the information content of realized Fall 2007 GPA. ε_{it+1} in equation (4.3) can be expressed as a function of surprises (new information) i.e.:

$$\varepsilon_{it+1} = h[\omega_{it+1} - E(\omega_{it+1}|\Omega_{it})] \tag{4.4}$$

Equation (4.3) can now be written as:

$$\Pr(X = 1|\Omega_{it+1}) - \Pr(X = 1|\Omega_{it}) = h[\omega_{it+1} - E(\omega_{it+1}|\Omega_{it})] \tag{4.5}$$

which basically states that the change in an individual's expectation between time t and $t+1$ about some event $\mathbf{X}$ that is realized at some point in the future is a function of surprises between time t and $t+1$. This equation highlights the challenges one faces in studying the updating of expectations; not only does the researcher need data on expectations of an agent over time but also needs to identify new information between periods. Bernheim (1988) uses assumptions on prior expectations in order to identify a model of revisions of Social Security benefit expectations. However, this defeats the purpose of collecting subjective expectations data. Dominitz (1997) faces the same problem in his analysis of revisions of earnings expectations in the SEE and, in the absence of knowledge of what the new information is, cannot pin down the causal explanation for the revision in expectations.

In order to come up with a metric of new information that wasn't anticipated at time t, I define ω_{it+1} to be the individual's GPA at the end of Spring 2007 (which is not known at time t but has been realized at time $t+1$). More specifically, ω_{it+1} equals 1 if i's cumulative GPA at the end of Spring 2007 was above the same threshold as was used for X (Spring 2008 GPA, or Graduating GPA). $E(\omega_{it+1}|\Omega_{it})$ is i's belief elicited in the Fall 2006 survey that $\Pr(\omega_{it+1}=1|\Omega_{it})$. Therefore, the metric $\omega_{it+1}-E(\omega_{it+1}|\Omega_{it})$ varies from -1 (this is the case of extreme negative surprise where the individual expected the Spring 2007 GPA to be above the threshold with certainty in the Fall 2006 survey but that did not happen) to 1 (in the case of extreme positive surprise). Panel A of Figure D.3 depicts the local linear estimates of equation (4.5), i.e. the regression of change in the Spring 2008 GPA beliefs on new information. Revisions of Spring 2008 GPA expectations seem to be positively related to the new information. Individuals who receive positive information increase their prediction of Spring 2008 GPA by less than a point-for-point increase. This makes sense since the dependent variable is a weighted index of one's performance in all quarters up to that point in time. Individuals who receive negative information only revise their predictions downward if the information content is very negative (less than -0.50).

Panel B of Figure D.3 estimates the regression function of equation (4.5) where the content of new information is defined as before, but $\mathbf{X}$ is now the GPA in one's major at the time of graduation. In contrast to revisions in Spring 2008 GPA beliefs, individuals do not revise their beliefs about their graduating GPA in response to the new information that is acquired between the two surveys. Since individuals have another year and a half of classes to take before this outcome is realized, it seems that they don't think that their Spring 2007 GPA gives them that much new information about their future performance, and that their expected cumulative GPA would still be the same. Moreover, Panel B shows that all individuals revise their beliefs about graduating GPA downwards, with those who do better than expected in Spring 2007 revising them downwards less. It should be pointed out that I only include the Spring 2007 GPA in ω_{it+1}. It could be the case that individuals are receiving some other information at the same time. However, as mentioned earlier, it is nearly impossible to identify all the new information. The analysis in this section is a preliminary attempt to understand how individuals revise expectations and, despite the very restrictive definition of the metric of surprises, individuals seem to respond in reasonable ways to new information.

4.4.2. Revisions of various major-specific beliefs

The discussion in section 4.4.1 highlights the breadth of data that is required to understand the revision of expectations in response to new information. Unfortunately, I don't have data for similar metrics of surprise for other determinants. However, one can still empirically investigate the evolution of beliefs. The underlying assumption is that individuals adopt a Bayesian learning approach. If the beliefs of the individuals can be characterized by a beta distribution (which is ideally suited to analyze binary events), the posterior probability P_{jm}^{t+1} (the probability of outcome j happening in the case of major

m) is given by:[16]

$$P_{jm}^{t+1} = \frac{\alpha}{\alpha+\beta} P_{jm}^{t} + \frac{\beta}{\alpha+\beta} I_{jm} \tag{4.6}$$

where P_{jm}^{t} is the prior belief of outcome j in major m, I_{jm} is the new information that the individual acquires about this outcome between period t and $t+1$, α is the precision of the prior, and β is the precision of the the new information. In the context of this study, the prior belief refers to the subjective belief elicited in the initial survey (Fall 2006), while the posterior refers to the belief elicited in the follow-up survey. The problem is that the researcher does not necessarily observe I_{jm}. Therefore, I use the following regression framework for the empirical investigation of (4.6):

$$P_{jm}^{t+1} = \gamma P_{jm}^{t} + \eta_{jm} + \varepsilon_{jm} \tag{4.7}$$

where ε_{jm} is a random error term, and:

$$\gamma = \frac{\alpha}{\alpha+\beta}; \quad \eta_{jm} = \frac{\beta}{\alpha+\beta} I_{jm}$$

The coefficients γ and η_{jm} show the nature of the learning process. One would expect γ to be equal to 1 and η_{jm} to be equal to 0 if the individual solely depends on her prior information, and does not learn any new information about the outcome between periods t and $t+1$. On the other hand, if the new information is really valuable, γ would be close to zero and η_{jm} would be large. Equation (4.7) is estimated for each of major-specific outcomes, and for three different majors in the individual's choice set. The results are

[16]see Viscusi and O'Connor (1984), and Viscusi (1997)

shown in Table D.14. The estimates are between the two extremes, and the prior belief continues to play a significant role in almost all the cases. Another object of interest is the importance of new information relative to the prior, which I denote as R. In the context of equation (4.6), $R = \frac{\beta}{\alpha}$. This is given as:

$$R = \frac{\beta}{\alpha} = \frac{1}{\gamma} - 1$$

The third row in each panel in Table D.14 shows the estimates of R. For outcomes like approval of parents, new information does not seem to be valuable ($R < 0.6$ in all cases). This is plausible because one would expect parents to have well-defined perceptions of different majors when a student starts college and, therefore, students are less likely to revise their beliefs about parents' approval over time. Similarly, priors for outcomes like graduating with a GPA of more than 3.5, and expected salary at the age of 30 are fairly precise. On the other hand, for outcomes related to the workplace, it seems that the prior is less precise; the metric R is, on average, larger for these outcomes. These findings are consistent with students adopting a Bayesian learning approach; for outcomes associated with college, one would expect the students to have fairly precise information at the time of the initial survey and, hence, for those outcomes, the relative importance of any new information will be less. Conversely, for outcomes in the workplace, one would expect students to receive *useful* information between the two surveys and, hence, the relative importance of new information would be higher in this case.

One would expect individuals to only revise the beliefs of outcomes of majors that they have *actively* thought about. To be more precise, an individual who was never interested in a major is less likely to revise her beliefs about the outcomes related to that major since she is not very likely to acquire any new information about it. The second column in Table D.14 shows the results for the major that was stated to be the individual's least preferred major at the time of the follow-up (and had been ranked in the lower half of one's preference ordering in the first survey). The results seem to support this theory-estimates of R for most outcomes are the lowest for the least preferred major.

4.4.3. Variance in beliefs

Another testable implication of the model outlined in section 4.4.1 is the change in cross-sectional variance of beliefs over time. From equation (4.3), one can write:

$$
\begin{aligned}
(4.8)\quad Var[\Pr(\mathbf{X} &= 1|\Omega_{it+1})] = Var[\Pr(\mathbf{X} = 1|\Omega_{it})] + Var[\varepsilon_{it+1}] \\
&= Var[\Pr(\mathbf{X} = 1|\Omega_{it})] + Var[\Pr(\mathbf{X} = 1|\Omega_{it+1}) - \Pr(\mathbf{X} = 1|\Omega_{it})] \\
&> Var[\Pr(\mathbf{X} = 1|\Omega_{it})]
\end{aligned}
$$

which implies that the variance of beliefs over time should increase. Here, I only test this implication on the labor force beliefs of the survey respondents. The lower panel of Table D.4 shows the mean beliefs of full-time labor force participation at the ages of 30 and 40 for males and females elicited in the two surveys. The variance in the beliefs for both genders and at both ages reported in the Fall 2007 survey is greater than the beliefs for the corresponding variables in the earlier survey. Therefore, this analysis lends support to the test that variance in beliefs increase over time.

4.5. Experimentation with majors

It is plausible that individuals learn about their ability and match quality in different majors by taking courses in them. One possible outcome of this is that the individual switches her major (see Manski, 1989, where he treats the decision of postsecondary schooling enrollment as an experiment, one possible outcome of which is dropout). Altonji (1993) outlines a simple theoretical model where individuals learn about their preferences between two fields of study by taking courses. Finally, Malamud (2006) develops a model in which individuals learn about their match quality by taking courses in them, and also gain information about match quality in the occupations specific to the field. Of the 117 survey respondents, 72 (~60%) report that they pursued (and then dropped) at least one other major in the past. More specifically, as depicted in Table D.15, 44 individuals reported to have experimented with one other major, while 28 stated that they had pursued at least two other majors.

In the context of the choice model in this study, individuals may learn about any of the major-specific outcomes by taking courses in it. A change in an individual's beliefs about her ability (graduating GPA, probability of completing the major in 4 years), match quality in college (outcomes like enjoying coursework), or match quality in workplace (enjoying working at the jobs, expected earnings at the jobs) could lead her to drop her current major. In order to understand the pattern of major switches, one would need data on the subjective beliefs about major-specific outcomes at several points in time over a short time horizon. In the absence of that data, I can, at best, only conduct a descriptive analysis of why individuals experiment with different majors.

Table D.16 shows some descriptive statistics of experimentation with majors. Individuals who experiment with only one major experience a small average gain of about 0.17 points in their GPA. Less than 50% of these individuals experience a positive change in their GPA indicating that academic performance is not the only dimension that influences one's choice of major. On average, individuals take about 3 courses in the major before dropping it. Respondents were asked to assign weights to different reasons for dropping the major such that they summed to a 100. Both males and females state losing interest in the original major, getting interested in something else, and the initial major being too challenging as the main causes for dropping the initial major. Table D.16 also shows these statistics for individuals who dropped more than one major. I show the results for the first two major switches. Individuals do experience a small net gain in GPA after switching their majors, but only about 40% of the individuals improve their GPA in the process. On average, individuals take 2.4 courses in the first major, and 1.81 courses in the second major. As in the case of single major drops, losing interest in the original major, getting interested in something else, and the initial major being too challenging are reported to be the main reasons for switching majors. Unfortunately, in the absence of detailed subjective data over time, it's not possible to say anything more about experimentation with majors.

4.6. Conclusion

One can infer how individuals choose college majors by eliciting individual's subjective expectations about major-specific outcomes and combining them with choice data (this is what was done in Zafar, 2007). However, in order to make credible policy recommendations, it is crucial to understand how individuals form expectations and revise them. In this paper, I undertake this task.

On the methodology side, the results in this paper bode well for the use of subjective expectations. I find that changes in expectations about various major-specific outcomes vary in sensible ways. Dominitz (1998) and Lochner (2007) reach similar conclusions but, unlike the current study, they only focus on a single variable; the former focuses on income expectations while the latter focuses on revisions to beliefs about arrest. I find that priors for outcomes like approval of parents, and graduating with a GPA of more than 3.5 are fairly precise, and that individuals don't revise them by as much as they revise their priors for outcomes that are realized in the workplace. Revisions of expectations of future GPA are positively related to changes in GPA between the two surveys. However, one needs some measure of new information (revealed between the two surveys) to understand how individuals revise expectations. I am able to come up with such a metric for GPA revisions by combining elicited expectations of GPA at various points in time with their realizations. I find that individuals only revise their short-term GPA in response to new information. Moreover, the change in beliefs about future GPA in response to positive and negative information is almost symmetric.

A second goal of this paper was to understand why females believe they won't enjoy studying certain fields, like Engineering and Math & Computer Science. I elicited various measures of monetary and non-monetary perceptions of discrimination. Both males and females seem to be aware of a positive wage gap in favor of males in most fields. However, they tend to underestimate the extent of the wage gap, and incorrectly believe that the wage gap stays roughly constant over time. Moreover, more males than females attribute the wage gap to "characteristics and aptitudes actually being different between males and females", while a larger fraction of females state "employers expecting different characteristics between males and females" as one of the main reasons for the wage gap. As shown in Filippin (2003), such a combination of beliefs can be self-confirming in a game-theoretical equilibrium and can cause females to make certain choices. The survey also elicited individuals' beliefs about males and females being *treated poorly* in the jobs in various majors. Perceptions of being treated poorly in the jobs in a given major are found to be correlated with the fraction of people of the same gender majoring in the field, the wage gap in the field, and beliefs about enjoying coursework and working at the jobs. The occupational segregation by gender has been documented by several studies in social psychology (Anker, 1997; Cejka and Eagly, 1999). However, it is not clear how to interpret the positive correlation between beliefs of females being treated poorly in the jobs and fraction of females taking classes in that major. The inclusion of this variable in a model of college major choice (initially estimated in Zafar, 2007) does not change the results since the variable is positively correlated with beliefs of enjoying coursework and enjoying working at the jobs.

There are at least two directions that can be taken from here. The first deals with the methodological aspect of this paper. As mentioned in section 4.4, identifying the information set of an individual is extremely challenging. In this paper, because of limited data that identifies new information, I only focus on GPA beliefs to study how individuals revise expectations. In order to enhance our understanding of expectations formation, it is crucial to collect repeated data on subjective expectations over a short time horizon.

Moreover, as argued in Manski (2004), rich longitudinal data on subjective expectations may not suffice to understand expectations formation, and probing people to learn how they perceive their environments may be informative.

From an applied aspect, it seems that individuals are forming their beliefs for various outcomes even before they come to college; for most major-specific outcomes, the prior belief continues to be important. In order to understand the gender gap in college major choice, it might be useful to focus on individuals in high school, and conduct a systematic analysis of their subjective beliefs. Finally, it might be useful to explore the role model hypothesis at lower levels of schooling.

References

[1] Abbink, K., B. Irelenbusch, and E. Renner, 2006. "Group Size and Social Ties in Microfinance Institutions", Economic Inquiry, 44: 614-628

[2] Alesina, A. and P. Giuliano (2007), "The Power of the Family", Harvard University Working Paper

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

[4] Altonji, J. and R. Blank (1999), "Race and Gender in the Labor Market", in O. Ashenfelter and D. Card, (eds), Handbook of Labor Economics Volume 3c Elsevier Science B.V.: 3144-3259

[5] American Association of University Women (AAUW) Educational Foundation (1998), "Separated by Sex: A Critical Look at Single-Sex Education for Girls", Washington, DC: AAUW

[6] Anderson, L., and C. Holt (1997), "Information Cascades in the Laboratory", American Economic Review, 87(5): 847-862

[7] Anderson, S., J. Goeree, and C. Holt (1998), "A Theoretical Analysis of Altruism and Decision Error in Public Goods Games", Journal of Public Economic 70(2): 297-323

[8] Andreoni, J. (1990), "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving", Economic Journal, 100(401): 464-77

[9] Andreoni, J., and B. Bernheim (2007), "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects", Working Paper.

[10] Andreoni, J., and R. Petrie (2004), "Public Goods Experiments without Confidentiality: A Glimpse into Fund-raising", Journal of Public Economics 88(7-8): 1605–1623

[11] Anker, R. (1997), "Theories of Occupational Segregation by Sex: An Overview", International Labour Review, 136(3): 315-339

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

[13] Asch, S. (1946), "Forming Impression of Personality", Journal of Abnormal Psychology, 41: 258-290

[14] Bamberger, G. (1986), "Occupation Choice: The Role of Undergraduate Education", Unpublished paper. Chicago: University of Chicago, Graduate School of Business

[15] Bandiera, O., I. Barankay, and I. Rasul (2007), "Social Incentives in the Workplace" LSE Working Paper

[16] Banerjee, A. (1992), "A Simple Model of Herd Behavior", The Quarterly Journal of Economics, 107(3): 797-817

[17] Bardsley, N. (2000), "Control without deception: Individual Behavior in Free-riding experiments revisited", Experimental Economics 3, 215-240

[18] Bardsley, N., and R. Sausgruber (2002), "Social interaction effects in the laboratory and society: conformism and reciprocity in public good provision", Working Paper Series. University of Nottingham, Nottingham.

[19] Baron-Cohen, S. (2003), "The Essential Difference: Men, Women, and the Extreme Male Brain", Allen Lane, London

[20] Barres, B. (2006), "Does Gender Matter?", Nature 442: 133-136

[21] Bernheim, B. (1988), "How Do Elderly Form Expectations? An Analysis of Responses to New Information." NBER Working Paper No. 2719

[22] Bernheim, B. (1994), "A Theory of Conformity", The Journal of Political Economy, 102(5): 841-877

[23] Bertrand, M. and S. Mullainathan (2001), "Do People Mean What They Say? Implications for Subjective Survey Data", American Economic Review Papers and Proceedings, 91(2): 67-72

[24] Bettinger, E. and B. Long (2005) "Do Faculty Server as Role Models? The Impact of Instructor Gender on Female Students,"American Economic Review Papers and Proceedings, 95(2): 152-157

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

[26] Betts, J. and D. Morell (1999), "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects", Journal of Human Resources 34(2): 268-293

[27] Bikhchandni, S., D. Hirshleifer, and I. Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", Journal of Political Economy, 100: 992-1026

[28] Bisin, A. and T. Verdier (2001), "The Economics of Cultural Transmission and the Dynamics of Preference", Journal of Economic Theory, 97: 298-319

[29] Bisin, A., G. Topa (2003), "Empirical Models of Cultural Transmission", Journal of the European Economic Association, 1(2-3): 363-375

[30] Bowles, S. and H. Gintis (2003), "Prosocial Emotions", Working Paper

[31] Brandts, J., and E. Fatas (2001), "Social Information and Social Influence in an experimental dilemma game", Economic Working Papers at Centro de Estudios Andaluces No 2004/59

[32] Breen R. and C. Garcia-Penalosa (2002), "Bayesian Learning and Gender Segregation", Journal of Labor Economics, 20(4): 899-922

[33] Brown, C. and M. Corcoran (1997), "Sex Based Differences in School Content and the Male-Female Wage Gap", Quarterly Journal of Economics, 99: 31-44

[34] Brownstone, D. and K. Train (1999), "Forecasting New Product Penetration with Flexible Substitution Patterns", Journal of Econometrics, 89: 109-129

[35] Bruine de Bruin, W., B. Fischhoff, B. Halpern-Felsher, and S. Millstein (2000), "Expressing Epistemic Uncertainty: It's a Fifty-Fifty Chance", Organizational Behavior and Human Decision Processes, 81: 115-131

[36] Canes, B. and H. Rosen (1995), "Following in Her Footsteps? Women's Choice of College Majors and Faculty Gender Composition", Industrial and Labor Relations Review, 48(3): 482-504

[37] Cason, T., and V. Mui (1998), "Social Influence in the sequential dictator game", Journal of Mathematical Psychology 42, 248-265

[38] Cejka, M, and A. Eagly (1999), "Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment", Personality and Social Psychology Bulletin, 25(4): 413-423

[39] Charness, G., and M. Dufwenberg (2006), "Promises and Partnership", Econometrica, 74(6):1579-1601

[40] Chemers, M., L. Hu, and B. Garcia (2001), "Academic Self-Efficacy and First-Year College Student Performance and Adjustment," Journal of Educational Psychology, XLIII: 55-64

[41] Cialdini, R. (1993), "Influence: Science and Practice", HarperCollins College

[42] Cooley, J. (2006), "Desegregation and the Achievement Gap: Do Diverse Peers Help", Working Paper

[43] Croson, R., and J. Shang (2005), "Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods", Wharton Working Paper

[44] Cross, S., and L. Madson (1997), "Models of the Self: Self-Construals and Gender", Psychological Bulletin, 122: 5-37

[45] De Giorgi, G., M. Pellizzari, and S. Redaelli (2007), "Be as Careful of the Books you Read as of the company you keep. Evidence on peer effects in educational choices", IGIER-Bocconi and IZA Working Paper.

[46] Delavande, A. (2004), "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice", Northwestern Working Paper

[47] Delavande, A. (2007), "Measuring Revisions to Subjective Expectations", Journal of Risk and Uncertainty, forthcoming.

[48] Deutsch, M., and B. Gerard (1955), "A Study of Normative and Informational Influences upon Individual Judgement", Journal of Abnormal and Social Psychology, 51: 629-636

[49] Dey, J. and C. Hill (2007), "Behind the Gender Gap", American Association of University Women Educational Foundation

[50] Di Tella, R., S. Galiani, and E. Schargrodsky (2007), "The Formation of Beliefs: Evidence From the Allocation of Land Titles to Squatters", The Quarterly Journal of Economics, 122 (1): 209-241

[51] Doepke, M. and F. Zilibotti (2007), "Occupational Choice and the Spirit of Capitalism," Quarterly Journal of Economics, forthcoming

[52] Dominitz J. (1998), "Earning Expectations, Revisions, and Realizations", Review of Economics and Statistics, 80: 374-388

[53] Dominitz J. and C. Manski (1996), "Eliciting Student Expectations of the Returns to Schooling", Journal of Human Resources, 31(1): 1-26

[54] Dominitz J. and C. Manski (1997), "Using Expectations Data to Study Subjective Income Distribution", Journal of the American Statistical Association, 87: 855-867

[55] Duncan, G., J. Boisjoly, M. Kremer, D. Levy, and J. Eccles (2005), "Peer Effects in Drug Use and Sex Among College Students", Journal of Abnormal Child Psychology, 33(3): 375-385

[56] Eagly, A. (1978), "Sex Differences in Influenceability", Psychological Bulletin, 85: 86-116

[57] Easterlin, R. (1995), "Preferences and Prices in Choice of Career: The Switch to Business, 1972-87", Journal of Economics Behavior and Organization, 27: 1-34

[58] Eide, E. and J. Grogger (1995), "Changes in College Skills and the Rise in the College Wage Premium", Journal of Human Resources 30: 280-310

[59] El-Gamal, M. A. and D. M. Grether (1995), "Are People Bayesian? Uncovering Behavioral Strategies", Journal of the American Statistical Association, 90 (432): 1137-1145

[60] Elster, J. (1999), "Alchemies of the Mind: Rationality and the Emotions", Cambridge University Press.

[61] Etzkowitz, H., C. Kemelgor, M. Neuschatz, and B. Uzzi (1992), "Athena Unbound: Barriers to Women in Academic Science and Engineering," Science and Public Policy, 19: 157-179

[62] Fairlie, R. (1999), "The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment", Journal of Labor Economics, 17(1): 80-108

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

[64] Fernandez, R. (2007a), "Culture and Economics", New Palgrave Dictionary of Economics, 2nd edition, forthcoming

[65] Fernandez, R. (2007b), "Culture as Learning: The Evolution of Female Labor Force Participation over a Century", NYU Working Paper

[66] Fernandez, R. and A. Fogli (2005), "Culture: An Empirical Investigation of Beliefs, Work, and Fertility", NBER Working Paper No. 11268

[67] Fernandez, R., A. Fogli, and C. Olivetti (2004), "Mothers and Sons: Preference Formation and Female Labor Force Dynamics", Quarterly Journal of Economics, 119(4): 249-299

[68] Filippin, A. (2003), "Discrimination and Workers' Expectations", IZA Discussion Paper n. 823

[69] Filippin, A., and A. Ichino (2005), "Gender wage gap in expectations and realizations", Journal of Labor Economics, 12(1): 125-145

[70] Fiorito, J., and R. Dauffenbach. (1982) "Market and Nonmarket Influences in Curriculum Choice by College Students", Industrial and Labor Relations Review, 36(1): 88-101

[71] Fischbacher, U. (2007), "z-Tree: Zurich Toolbox for Ready-made Economic Experiments:, Experimental Economics 10(2): 171-178

[72] Fischbacher, U., S. Gachter, and E. Fehr (2001), "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", Economics Letters, 71(3): 397-404

[73] Fischhoff, B., A. Parker, W. de Bruin, J. Downs, C. Palmgren, R. Dawes, and C. Manski (2000), "Teen Expectations for Significant Life Events", Public Opinion Quarterly, 64(2): 189-205

[74] Fogli, A., and L. Veldkamp (2007), "Nature or Nurture? Learning and Female Labor Force Dynamics", Working Paper

[75] Freeman, R. (1971) "The Market for College-Trained Manpower", Cambridge, Mass.: Harvard University Press

[76] Frey, B, and S. Meier (2004), "Social Comparison and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment", American Economic Review 94(5): 1717-1722

[77] Garman, D. and L. Loury (1995), "College Selectivity and Earnings", Journal of Labor Economics, 13: 289-308

[78] Goeree, J., and L. Yaariv (2006), "Conformity in the Lab", Caltech Working Paper.

[79] Goldin, C., L. Katz, and I. Kuziemko (2006), "The Homecoming of American College Women: The Reversal of the College Gender Gap", Journal of Economic Perspectives, 20(4): 133-156

[80] Gruber, J. (2004), "Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce", Journal of Labor Economics, 22(4): 799-833

[81] Guiso, L., P. Sapienza, and L. Zingales (2006), "Does Culture Affect Economic Outcomes?", The Journal of Economic Perspectives, 20(2): 23-48

[82] Haan, M., P. Kooreman, and T. Riemersma (2006), "Friendship in a Public Good Experiment", IZA Discussion Paper No. 2108

[83] Harbaugh, W. (1998a), "The Prestige Motive for Making Charitable Transfers", American Economic Review, 88(2): 277-282

[84] Harbaugh, W. (1998b), "What do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow", Journal of Public Economics, 67: 269-284

[85] Houser, D., M. Keane and K. McCabe (2004), "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm", Econometrica, 72(3): 781-822

[86] Hurd, M., and K. McGarry (2002), "The Predictive Validity of Subjective Probabilities of Survival," Economic Journal, 112(482): 966-985

[87] Jones, S. (1984), "The Economics of Conformism", Oxford: Basil Blackwell

[88] Keser, C., and F. van Winden (2000), "Conditional cooperation and voluntary contributions to public goods",Scandinavian Journal of Economics 102: 23-39

[89] Kimura, D. (1999), "Sex and Cognition", Cambridge, Mass: A Bradford Book, MIT Press

[90] Kimura, D. (2006), ""Under-representation" or Misrepresentation?", in Why Aren't More Women In Science? S.J. Ceci & W. Williams (Eds.), APA Books: 39-46

[91] Kohler, H., J. Behrman, and S. Watkins (2001), "The Density of Social Networks and Fertility Decisions: Evidence from S. Nyanza District, Kenya". Demography, 38(1), 43-58

[92] Lafortune, J. (2008), "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market", MIT Working Paper

[93] Landry, C., A. Lange, J. List, M. Price, and N. Rupp (2006), "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment," Quarterly Journal of Economics,121 (2): 747-782

[94] Leach, E. (1982), "Social Anthropology", Glasgow: Fontana.

[95] Ledyard, J. (1995), "Public goods: a survey of experimental research", in Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ.

[96] Light, A., and M. Ureta (1995), "Early-Career Work Experience and Gender Wage Differentials", Journal of Labor Economics, 13(1): 121-154

[97] Lochner, L. (2007), "Individual Perceptions of the Criminal Justice System", American Economic Review, 97(1), 444-460

[98] Long, V. (1986), "Relationship of Masculinity to Self-Esteem and Self-Acceptance in Female Professionals, College Students, Clients, and Victims of Domestic Violence", Journal of Consulting and Clinical Psychology, 54: 323-327

[99] Loprest, P. (1992), "Gender Differences in Wage Growth and Job Mobility", American Economic Review, 82(2): 526-532

[100] Lubinski, D., and C. Benbow (1992), "Gender differences in abilities and preferences among the gifted: Implications for the math/science pipeline", Current Directions in Psychological Science, 1: 61-66

[101] Luce, D. and P. Suppes (1965), "Preferences, utility and subjective probability'", in R. Luce, R. Bush, and E. Galanter, eds., Handbook of Mathematical Psychology, John Wiley and Sons, New York: 249–410

[102] Malamud, O. (2006), "Breadth vs. Depth: The Effect of Academic Specialization on Labor Market Outcomes", Harris School Working Paper Series 05:17

[103] Malamud, O. (2006), "Breadth vs. Depth: The Effect of Academic Specialization on Labor Market Outcomes", Harris School Working Paper Series 05:17

[104] Mandell, R. (N. D). "Evaluating Short Versions of the Marlowe-Crowne Social Desirability Scale", HTML document. http://www.od.nih.gov/behaviorchange/measures/PDF/marlowe-crowne.pdf.

[105] Manski, C. (1989), "Schooling as Experimentation: a Reappraisal of the Postsecondary Dropout Phenomenon", Economics of Education Review, 8(4): 305-312

[106] Manski, C. (1993a), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in C. Clotfelter and M.Rothschild (editors) Studies of Supply and Demand in Higher Education, Chicago: University of Chicago Press: 43-57

[107] Manski, C. (1993b), "Identification of Endogenous Social Effects: The Reflection Problem", Review of Economic Studies, 60: 531-542

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

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

[110] Manski, C. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", Review of Economic Studies, 60: 531-542

[111] Manski, C. (2000), "Economic Analysis of Social Interactions", Journal of Economic Perspectives, 14(3): 115-136

[112] Manski, C., and F. Molinari (2008), "Rounding Probabilistic Expectations in Surveys", Northwestern Working Paper

[113] Mas, A., and E. Moretti (2006), "Peers at Work", mimeo, University of California, Berkeley

[114] McLanahan, Sara, and Gray Sandefur (1994), "Growing up with a Single Parent", Cambridge, MA: Harvard University Press

[115] Messick, D., (1999), "Alternative Logics for Decision Making in Social Settings", Journal of Economic Behavior and Organization 39(1): 11-28

[116] Messick, D., H. Wilke, M. Brewer, and R. Kramer, 1983. "Individual Adaptations and Structural Change as Solutions to Social Dilemmas", Journal of Personality and Social Psychology, 44(2): 294-309

[117] Moffitt, R. (2001), "Policy Interventions, Low-Level Equilibria, and Social Interactions", in Durlauf S., and H. Young (eds.) Social Dynamics, Washington, DC: Brookings Institution Press

[118] Munshi, K., and J. Myaux (2006), "Social Norms and the Fertility Transition", Journal of Development Economics 80(1):1-38

[119] Niederle, M. and L. Vesterlund (2007), "Do Women Shy away from Competition? Do Men Compete too Much?", Quarterly Journal of Economics, 122(3): 1067-1101

[120] Oaxaca, R. (1973), "Male-Female Wage Differential in the Urban Labor Markets", International Economic Review, 14(October): 693-709

[121] Ouchi, W. (1981), "THEORY Z: How American Management Can Meet the Japanese Challenge", Addison-Wesley

[122] Paglin, M. and A. Rufolo (1990), "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences", Journal of Labor Economics, 8: 123-144

[123] Polachek, S. (1978), "Sex Differences in College Major", Industrial and Labor Relations Review, 31(4): 498-508

[124] Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", American Economic Review 83, 1281-1302

[125] Rege, M., and K. Telle (2004), "The Impact of Social Approval and Framing on Cooperation in Public Good Situations", Journal of Public Economics, 88 (7-8): 1625-1644

[126] Reuben, E. and F. van Winden (2006), "Social Ties and Coordination on Negative Reciprocity: The Role of Affect", Journal of Public Economics, forthcoming

[127] Revelt, D. and K. Train (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", The Review of Economics and Statistics, LXXX (4): 647-657

[128] Sacerdote, B. (2001), "Peer Effects With Random Assignment: Results for Dartmouth Roommates", Quarterly Journal of Economics, 116 (May): 681-704

[129] Smith, H. and B. Powell (1990), "Great Expectations: Variations in Income Expectations Among College Seniors", Sociology of Education, 63 (July): 194-207

[130] Solnick, S. (1995), "Changes in Women's College Majors from Entrance to Graduation at Women's and Coeducational Colleges", Industrial and Labor Relations Review, 48(3): 505-514

[131] Sweeting, A. (2006), "Coordination, Differentiation and the Timing of Radio Commercials", Journal of Economics and Management Strategy, 15(4): 909-942

[132] Traweek, S. (1988), "Beamtimes and Lifetimes: The World of High Energy Physicists", Cambridge, MA: Harvard University Press

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

[134] Valian, V. (1998), "Why So Slow?", MIT Press, Cambridge.

[135] Van der Klaauw, W. (2000), "On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices", UNC-Chapel Hill Working Paper

[136] Vesterlund, L. (2003), "The Informational Value of Sequential Fundraising", Journal of Public Economics, 87(3-4): 627-657

[137] Vincent, P., L. Peplau, and C. Hill (1998), "A Longitudinal Application of the Theory of Reasoned Action to Women's Career Behavior", Journal of Applied Social Psychology, 28: 761-778

[138] Viscusi, W. (1997), "Alarmist Decisions with Divergent Risk Information", The Economic Journal, 107(445): 1657-1670

[139] Viscusi, W., and C. O'Connor (1984), "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?", American Economic Review, 74(5): 942-956

[140] Xie, Y. and K. Shauman (2003), "Women in Science: Career Processes and Outcomes", Cambridge, MA: Harvard University Press

[141] Yun, M. (2004), "Decomposing Differences in the First Moment", Economic Letters 82(2): 275-280

[142] Zafar, B. (2007), "College Major Choice and the Gender Gap", Northwestern Working Paper

APPENDIX A

Appendix 1 for Chapter 2

A.1. Survey Questions

A.1.1. Practice Questions

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

We will start with a couple of practice questions.

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

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

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

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

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

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

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

A.1.2. Questionnaire

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

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

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

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

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

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

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

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

Q7 If you were majoring in Natural Sciences, what do you think is the percent chance that you could find a job (that you would accept) immediately upon graduation?

Q8 If you obtained a bachelors in Natural Sciences, what do you think is the percent chance that you will go to graduate school in Natural Sciences some time in the future?

Q9 What do you think was the average annual starting salary of Northwestern graduates (of 2006) with Bachelor's Degrees in Natural Sciences?

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

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

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

Q10 What kind of jobs are you thinking of?

Q11 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, what do you think is the percent chance that you will enjoy working at the kinds of jobs that will be available to you?

Q12 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, what do you think is the percent chance that you will be able to reconcile work and your social life/ family at the kinds of jobs that will be available to you?

Q13 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, how many hours per week on average do you think you will need to spend working at the kinds of jobs that will be available to you?

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

Q14 Look ahead to when you will be 30 years old. Think about the kinds of jobs that will be available to you and that you will accept if you graduate in [X]. 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?

A.2. Debriefing

A.2.1. Why Choose Two Majors

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

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

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

A.2.2. Peer Effects

The question was:

Check all that apply

1) My (intended) major is the same as that of one of my parents

2) My (intended) major is the same as that of one of my siblings

3) My (intended) major is the same as that of my freshman-year roommate

4) My (intended) major is the same as that of my current roommate

5) My (intended) major is the same as that of the majority of my best high school friends who went to college

6) My (intended) major is the same as that of the majority of my friends in Northwestern

7) None of the above

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

Some of the selected responses are:

- I am influenced by my father but not much by friends.
- My Integrated science major is the same as the majority of my friends, because most of the classes that I take is with Integrated science majors. Since we are in class together all the time, we have become good friends.
- My brother is majoring in Journalism but also Political Science. This played a minor influence on my decision but is mostly coincidence that we like the same sort of classes. My freshman year roommate was possibly an influence on me, but we generally had the same interests in terms of school subjects from the start.
- My dad majored in English, is passionate about the subject and is now a college professor who teaches it. He loved it, but it was never forced on me, resulting in that i grew to love it as well. And I'm good at it. When you're constantly being grammatically corrected and pushed to think loftier ideas then it kind of becomes

second nature, a permanent habit. As far as my freshman year roommate, i lived in the Communications Residential College. It's 80% journalism and 19% theater. It was bound to happen.

- My brothers and I have very similar interests and strengths.
- My parents have always encouraged me to do well in school, and placed an emphasis on math and the sciences. Also, I live in a town of only 20,000 people, but there are two major research facilities in the town. Many of my peers were also children of scientists. I have a twin brother who also goes to Northwestern and studies Chemistry and German. We probably influenced each other because we're very close. We both took the German AP, which is why both of us have German as a second major (the German major is relatively light, especially if you come in already taking third year classes).
- I am interested in Psychology, and although my parents are not too keen on me studying psychology, that's what I want to to. My mom was also interested in Psych, but she never perused it
- My major is the same as my parents purely by coincidence. Somehow our interests coincide. My major is the same as the majority of my high school friends (but most of my best friends are doing medicine) because most of my high school friends who study abroad chose economics. It is also the major which most students from Hong Kong would choose when they study abroad since most jobs you can find back home is econ-related. My major is the same as the majority of my friends in Northwestern because 1) Economics is a popular major, the probability that you can find an econ major student is quite high 2) I met

most of my friends and formed the friendship through classes and extracurricular activities.

- My mom is a psychologist, and even though I have no desire to pursue that career I think she might have influenced my interest in psychology
- I grew up in a household where my parents are both scientists so I became interested in medicine and science simultaneously. They never told me what to do, it was just a matter of spending more time around a certain field. Also, I live on North Campus where a majority of Northwestern science majors and engineers live so it just so happens that many people are in the same field that I intend to be in, primarily by location because the dorms up North are closer to Tech, which is where most of our classes are held.
- 1) Parental Influence 5 and 6) Social Integration with Friends of Similar Background
- For the first, my parents raised me and my siblings, and for the second, I tend to make friends with people I share classes with.
- I think they paired me with a roommate with whom I had stuff in common. My friends at Northwestern and I have the same interests and personalities and that is reflected in our majors.
- My roommate took a Psychology class last year and really enjoyed it. I had never had any exposure to Psychology classes in high school, so decided that it would be interesting to take. I took the class this fall, and really enjoyed it.
- My parents and I have similar tastes and I like the things they like. My roommate and I were best friends from high school and had very similar interests.

- I think I want to major in economics because I see how successful my dad is today and since he majored in business, I thought economics would be close enough.
- economics is something that flows for me when i learn it, maybe it's in my genes since my dad majored in it during graduate school, it's also very practical and covers many bases, so i see why my friends picked it, it's respected, it's not seen as a slacker major like psychology, and i find it very interesting as i would hope many people do since it's such a popular major
- I really think it's a coincidence. My roommate is interested in politics, too. Maybe it's because we're from similar places. We're both from coastal cities, where politics is big.
- My father has influenced me indirectly because he is an economics professor. My brother is young and wants to follow me into business. i am friends with a lot of people in my classes, which happen to be econ./MMSS classes
- My mother is terrible at math so she majored in an all-words major, Sociology, but I am OK at math so my Social Policy major incorporates a bit more economic reasoning and logic than hers

A.3. Figures and Tables

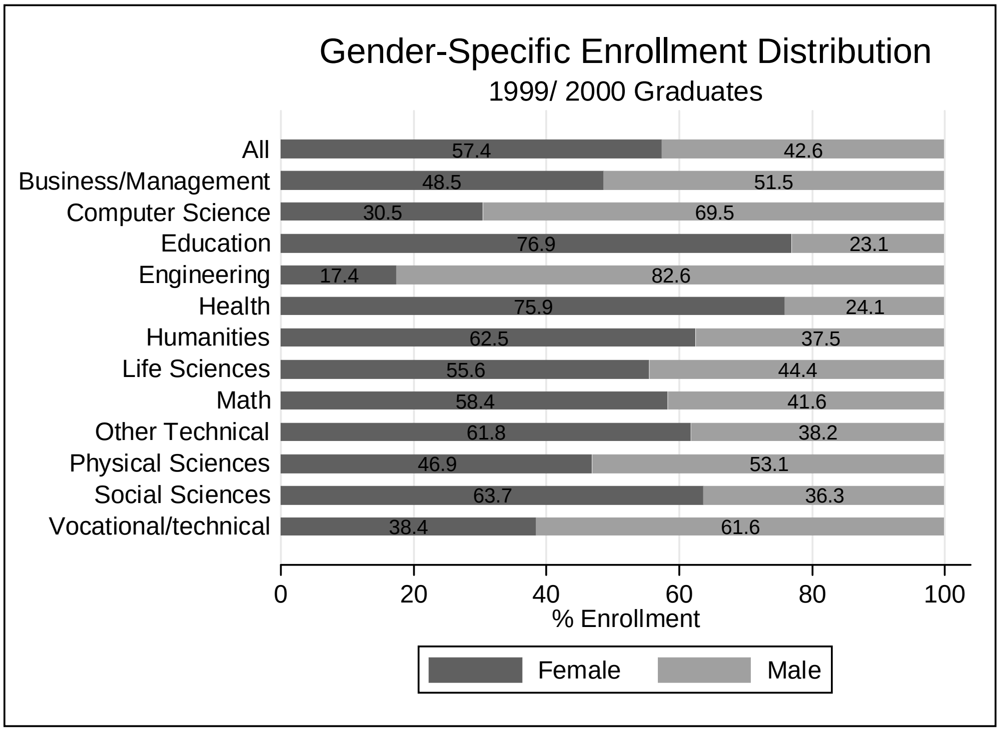

Figure A.1. Gender Composition of Undergrad Majors of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001.

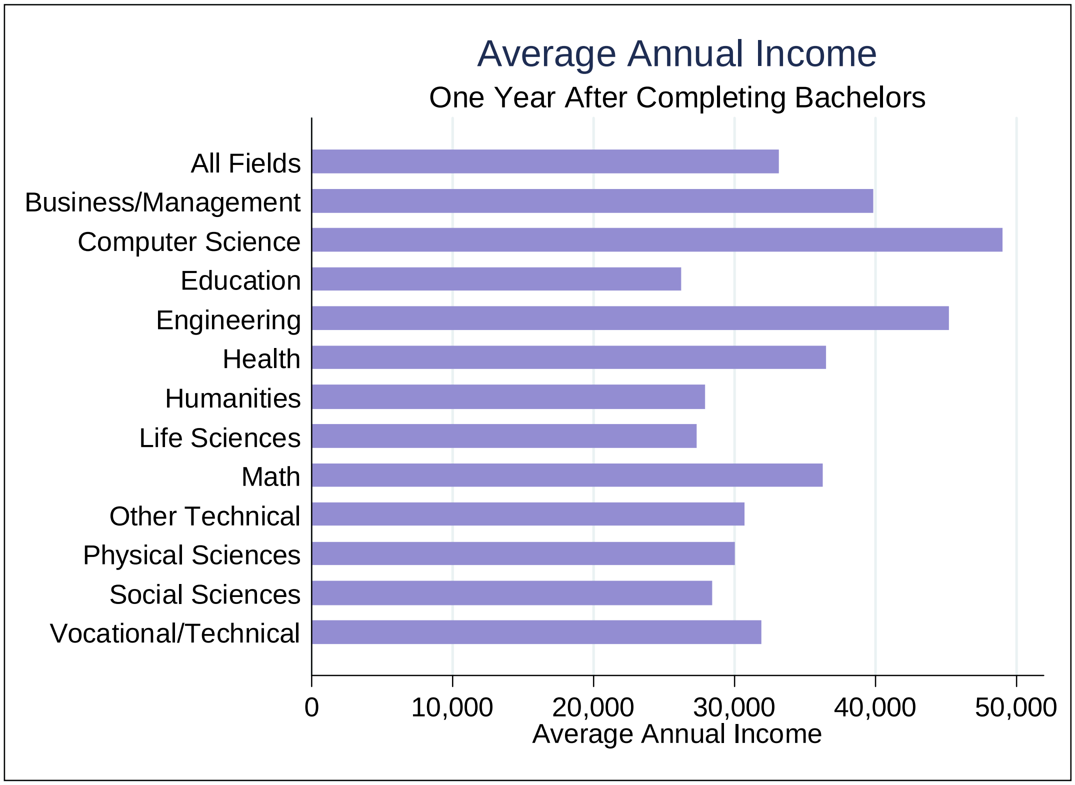

Figure A.2. Average income of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001 by Undergraduate Major.

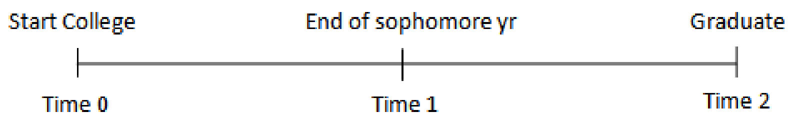

Figure A.3. Timeline

Table A.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 the Social Sciences*

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

f Area Studies
African American Studies
American Studies
Asian And Middle East Languages and 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 A.2. Sample Characteristics

	Sample			Population[a]
	All	Single Majors	Double Majors	
Characteristics	**Freq.(Percent)**	**Freq.(Percent)**	**Freq.(Percent)**	**Freq.(Percent)**
Gender				
Male	69 (43)	33 (40)	36 (46)	465 (46)
Female	92 (57)	50 (60)	42 (54)	546 (54)
Total	161	83	78	1011
Ethnicity				
Caucasian	79 (49)	40 (48)	39 (50)	546 (54)
African American	11 (7)	7 (8.5)	4 (5)	71 (7)
Asian	56 (35)	27 (33)	29 (37)	232 (23)
Hispanic	5 (3)	2 (2)	3 (4)	61 (6)
Other	10 (6)	7 (8.5)	3 (4)	101 (10)
Declared Major?[b]				
Yes	90 (56)	44 (53)	46 (59)	182 (18)
No	71 (44)	39 (47)	32 (41)	829 (82)
International Std?[c]				
Yes	8 (5)	5 (6)	3 (4)	40 (4)
No	153 (95)	78 (94)	75 (96)	971 (96)
Second-Gen Imm?[d]				
Yes	66 (41)	33 (40)	33 (42)	–
No	95 (59)	50 (60)	45 (58)	–
Average GPA				
Male	3.48	3.43	3.52	3.26
Female	3.40	3.39	3.45	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 is an international student
d Whether at least one of the respondent's parents is foreign-born, and the respondent was born in the US

Table A.3. Distribution of WCAS Majors

	Sample[b]						Class of 2006[c]					
	All		Males		Females		All		Males		Females	
WCAS Majors[a]	**Freq**	**(%)**	**Freq**	**(%)**	**Freq**	**(%)**	**Freq**	**(%)**	**Freq**	**(%)**	**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 1a
b In cases where the survey respondent has more than one major in WCAS, only the first one is included
c Only includes students with a primary WCAS major (Source: Integrated Postsecondary Education Data System)

Table A.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 A.5. Percent Chance of reconciling work and family at the jobs if majoring in:

	Social Sciences II				Social Sciences I			
	Males		Females		Males		Females	
Sub. Beliefs	**Freq.**	**Cum. %**	**Freq.**	**Cum. %**	**Freq.**	**Cum. %**	**Freq.**	**Cum. %**
0	1	1.45	-	0	1	1.45	-	0
10	-	1.45	2	2.17	-	1.45	1	1.09
12	-	1.45	1	3.26	-	1.45	-	1.09
20	-	1.45	2	5.43	-	1.45	-	1.09
25	1	2.90	1	6.52	1	2.90	-	1.09
30	2	5.80	3	9.78	-	2.90	1	2.17
35	1	7.25	2	11.96	-	2.90	-	2.17
40	3	10.14	8	20.65	1	4.35	3	5.43
43	-	10.14	1	21.74	-	4.35	-	5.43
44	1	11.59	-	21.74	-	4.35	-	5.43
45	3	15.94	3	22.83	-	4.35	1	6.52
46	-	15.94	-	23.91	-	4.35	-	6.52
47	1	17.39	-	23.91	-	4.35	-	6.52
50	8	28.99	10	34.78	6	13.04	4	10.87
55	-	28.99	5	40.22	1	14.49	1	11.96
56	1	30.43	1	41.30	-	14.49	-	11.96
60	9	43.38	9	51.09	5	21.74	9	21.74
65	3	47.83	3	54.35	2	24.64	2	23.91
68	1	49.28	-	54.35	-	24.64	-	23.91
70	3	53.62	8	63.04	7	34.78	14	39.13
71	-	53.62	-	63.04	-	34.78	1	40.22
72	-	53.62	1	64.13	1	36.23	-	40.22
74	-	53.62	1	65.22	-	36.23	-	40.22
75	13	72.46	9	75.00	7	46.38	6	46.74
76	1	73.91	-	75.00	1	47.83	2	48.91
77	-	73.91	-	75.00	1	49.28	-	48.91
78	-	73.91	-	75.00	1	50.72	-	48.91
79	-	73.91	-	75.00	-	50.72	-	48.91
80	10	88.41	7	82.61	11	66.67	18	68.48
82	-	88.41	1	83.70	-	66.67	2	70.65
83	-	88.41	1	84.78	-	66.67	-	70.65
85	3	92.75	5	90.22	4	72.46	7	79.26
86	1	94.20	-	90.22	-	72.46	-	79.26
87	-	94.20	-	90.22	-	72.46	1	79.35
88	-	94.20	-	90.22	-	72.46	1	80.43
89	-	94.20	1	91.30	-	72.46	1	81.52
90	2	97.10	5	96.74	12	89.86	10	92.39
93	-	97.10	-	96.74	1	91.30	1	93.48
95	1	98.55	2	98.91	1	92.75	3	96.74
96	-	98.55	-	98.91	-	92.75	1	97.83
97	1	100	-	98.91	-	92.75	-	97.83
98	-	100	1	100	2	95.65	1	98.91
99	-	100	-	100	-	95.65	1	100
100	-	100	-	100	3	100	-	100
Total	69		84		69		92	

Table A.6. Statistics on Average Annual Starting Salaries of 2006 Graduates

Variable: Average Annual Starting Salary of 2006 Northwestern Graduates in each Category as reported by:†											
	Graduating Class of '06[a]			*Major in the category*[b]				*Major not in the Category*[c]			
Category	**All**	**Males**	**Females**	**Males**		**Females**		**Males**		**Females**	
	(1)	(2)	(3)	4*a*	4*b*	5*a*	5*b*	6*a*	6*b*	7*a*	7*b*
	Avg.	Avg.	Avg.	Avg.	Median	Avg.	Median	Avg.	Median	Avg.	Median
Natural Sciences	38.56*	55.00	33.08	43.67	40.00	56.29	35.00	46.53	44.00	49.93	50.00
	[16][d]	[4]	[12]	[15]		[17]		[53]		(75)	
Math & Computer Sc	59.95	65.28	47.50	51.40	52.00	32.50	32.50	50.13	50.00	47.37	45.00
	[10]	[7]	[3]	[5]		[2]		[64]		[90]	
Social Sciences I	39.66	42.42	37.86	40.73	40.00	46.28	40.00	41.13	40.00	34.15	32.50
	[81]	[32]	[49]	[20]		[36]		[49]		[56]	
Social Sciences II	52.85*	55.07	49.31	58.08	55.00	62.77	57.50	60.06	50.00	50.93	50.00
	[96]	[59]	[37]	[38]		[22]		[31]		[70]	
Ethics and Values	37.88	35.50	45.00	32.60	35.00	34.50	34.50	39.11	35.00	36.14	31.00
	[4]	[3]	[1]	[5]		[2]		[64]		[90]	
Area Studies	38.25	40.50	35.00	59.00	46.00	35.14	36.00	36.72	35.00	34.16	31.00
	(4)	(3)	(1)	[8]		[22]		[61]		[70]	
Literature & Fine Arts	37.67	48.33	35.53	27.60	25.00	36.48	37.50	36.23	31.00	30.47	30.00
	[18]	[3]	[15]	[5]		[20]		[64]		[72]	
Music Studies	37.89	35.80	40.50	25.00	25.00	25.00	25.00	–		–	
	[9]	[5]	[4]	[1]		[2]					
Educ & Social Policy	43.50*	67.00	35.67	–		35.00	35.00	–		–	
	[16]	[4]	[12]			[2]					
Communication Std	42.53	50.83	38.51	–		43.33	40.00	–		–	
	[46]	[15]	[31]			[3]					
Engineering	53.49	54.53	52.55	50.00	52.50	50.00	50.00	53.19	50.00	55.55	50.00
	[109]	[71]	[38]	[4]		[1]		[63]		[83]	
Journalism	42.94*	80.12	36.09	32.50	32.50	40.00	40.00	–		–	
	[38]	[7]	[31]	[2]		[1]					

† Response to: What do you think was the average annual starting salary of Northwestern graduates (of 2006) with Bachelor's Degrees in Category X?
* gender difference significant (p-value < 0.05; two-tailed t-test); responses are in thousands.
[a] Avg reported starting salary of a graduating member of 2006 majoring in that category. (Northwestern University Graduation Survey 2006)
[b] Respondent's answer to † when (one of) his/her intended major is in category X. Column (a) gives the mean, and Column (b) gives the median
[c] Respondent's answer to the question in † when his/her intended majors are in a category other than X.
[d] [N]: N is the number of observations for the given statistic

Table A.7. Explaining the errors in students' salary expectations

Dependent Variable: Log Absolute Error in Beliefs about Starting Salaries[⊕a]	*Entire Sample*		*Overestimate*[†]		*Underestimate*[‡]	
	Estimates	(Std. Error)	Estimates	(Std. Error)	Estimates	(Std. Error)
	(1*a*)	(1*b*)	(2*a*)	(2*b*)	(3*a*)	(3*c*)
Major Declared[b]	−0.002	(0.106)	0.287	(0.193)	−0.144	(0.094)
Cumulative GPA	0.302**	(0.138)	0.486**	(0.243)	0.205*	(0.118)
SAT Math[c]	−0.0313	(0.0472)	−.0692	(0.0924)	−0.0065	(0.0389)
SAT Verbal[d]	-0.0186	(0.0404)	−0.088	(0.083)	−0.0207	(0.0349)
Female	0.199*	(0.113)	0.143	(0.225)	0.221**	(0.099)
NU Credits[e]	−0.0212	(0.0177)	0.0902***	(0.029)	0.0044	(0.0138)
Asian	0.0380	(0.168)	0.0301	(0.346)	0.0567	(0.127)
Foreign[f]	0.0525	(0.287)	−0.201	(0.696)	0.268*	(0.146)
Second-Generation Imm.[g]	0.142	(0.157)	0.285	(0.325)	0.0646	(0.103)
Studying Given Major[h]	−0.0034	(0.118)	0.0701	(0.246)	0.0145	(0.084)
Studying Major × Major Dec.	−0.186	(0.161)	−0.158	(0.291)	−0.0992	(0.119)
Private High School	0.0199	(0.104)	0.112	(0.179)	0.0017	(0.103)
Low Parents' Income[i]	−0.152	(0.115)	−0.274	(0.189)	0.0112	(0.102)
Father went to College	−0.124	(0.252)	−0.219	(0.548)	−0.142	(0.169)
Mother went to College	−0.247	(0.254)	−0.697	(0.438)	0.0259	(0.126)
Father studied major[j]	−0.114	(0.105)	−0.241	(0.173)	0.0106	(0.089)
Mother studied major[k]	0.057	(0.110)	0.064	(0.185)	0.0467	(0.0965)
Respondent Random Effect	Yes		Yes		Yes	
No. of Observations	1288		557		731	
No. of Clusters	161		128		141	

see notes on next page

Notes for Table A.7

Parameter estimates correspond to the estimation of OLS model. Cluster errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

$\oplus$Dependent variable is log of the absolute value of the salary error: $\ln|\frac{(\widehat{s_{ix}}-s^{obs}}{s^{obs}}|$ where $\widehat{s_{ix}}$ is respondent's answer to: "What do you think was the average annual starting salary of Northwestern graduates (of 2006) with Bachelor's Degrees in Category X?" s^{obs} is the actual average salary earned by 2006 graduates in category X (source: Northwestern Career Center Survey of 2006 Graduates)

† Sample restricted to observations where reported estimate is greater than observed salary, i.e. $\widehat{s_{im}} > s^{obs}$
‡ Sample restricted to cases where reported estimate is less than observed salary, i.e. $\widehat{s_{im}} < s^{obs}$
a All regressions include major-specific dummies, and respondent fixed effects. (Constants not shown)
b a dummy variable that equals one if the respondent has already declared his/ her major
c (*d*) - SAT Math (Verbal) score; = 1 if SAT Math (Verbal) score is less than 400; =2 if score = 400-499; =3 if score = 500-549; =4 if score = 550-599; =5 if score = 600-649; =6 if score = 650-699; =7 if score =700-749; =8 if score=750-800
e Number of credits the respondent gets when starting Northwestern because of AP/ IB exams
f a dummy that equals one if the respondent is an International student
g equals 1 if either of the respondent's parents are foreign-born, & respondent was born in the US
h equals 1 if respondent's intended major category is same as category X in the salary question
i a dummy that equals one if parents' annual income is less that $150,000
j a dummy that equals one if father's field of study is the same as the salary question
k a dummy that equals one if mother's field of study is the same as the salary question

Table A.8. Expected Annual Salary at the Age of 30

Variable: Avg. Annual Starting Salary of 2006 Northwestern Graduates in each 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.	(3*a*) Avg.	(3*b*) Med.	(3*c*) N^d	(4*a*) Avg.	(4*b*) Med.	(4*c*) N	(5*a*) Avg.	(5*b*) Med.	(5*c*) N	(6*a*) Avg.	(6*b*) Med.	(6*c*) 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 A.9. Percentage Chance of being active in the full-time labor force

	At the age of 30†						*At the ago of 40*‡					
Sub.	**All**		**Females**		**Males**		**All**		**Females**		**Males**	
Beliefs	**Freq**	**Cum%**	**Freq**	**Cum%**	**Freq**	**Cum%**	**Freq**	**Cum%**	**Freq**	**Cum%**	**Freq**	**Cum%**
30	-	0	-	0	-	0	2	1.24	2	2.17	-	0
35	-	0	-	0	-	0	1	1.86	1	3.26	-	0
40	1	0.62	1	1.09	-	0	2	3.11	2	5.43	-	0
45	1	1.24	1	2.17	-	0	-	3.11	-	5.43	-	0
50	3	3.11	3	5.43	-	0	3	4.97	3	8.70	-	0
55	1	3.73	1	6.52	-	0						
60	1	4.35	1	7.61	-	0	4	7.45	3	11.96	1	1.45
65	2	5.59	2	9.78	-	0	1	8.07	-	11.96	1	2.90
70	2	6.83	1	10.87	1	1.45	6	11.80	6	18.48	-	2.90
75	3	8.70	3	14.13	-	1.45	5	14.91	5	23.91	-	2.90
80	11	15.53	9	23.91	2	4.35	10	21.12	7	31.52	3	7.25
82	4	18.01	4	28.26	-	4.35	2	22.36	1	32.61	1	8.70
84	1	18.63	-	28.26	1	5.80	1	22.98	1	33.70	-	8.70
85	5	21.74	5	33.70	-	5.80	10	29.19	8	42.39	2	11.59
86	-	21.74	-	33.70	-	5.80	2	30.43	-	42.39	2	14.49
87	-	21.74	-	33.70	-	5.80	1	31.06	1	43.48	-	14.49
88	-	21.74	-	33.70	-	5.80	1	31.68	1	44.57	-	14.49
89	2	22.98	1	34.78	1	7.25	1	32.30	-	44.57	1	15.94
90	29	40.99	20	56.52	9	20.29	28	49.69	14	59.78	14	36.23
92	1	41.61	-	56.52	1	21.74	2	50.93	-	59.78	2	39.13
93	2	42.86	2	58.70	-	21.74	1	51.55	-	59.78	1	40.58
95	37	65.84	13	72.83	24	56.52	31	70.81	15	76.09	16	63.77
96	-	65.84	-	72.83	-	56.52	1	71.43	-	76.09	1	65.22
97	2	67.08	1	73.91	1	57.97	4	73.91	3	79.35	1	66.67
98	17	77.64	11	85.87	11	66.67	14	82.61	8	88.04	6	75.36
99	8	82.61	2	88.04	2	75.36	4	85.09	1	89.13	3	79.71
100	28	100.00	11	100.00	11	100.00	24	100.00	10	100.00	14	100.00
Total	161		92		69		161		92		69	

† This is the response to: "What do you think is the percent chance that you will be active in the FULL-TIME labor forcewhen you are 30 YEARS OLD? (i.e you will be working full-time at the age of 30) ASSUME that you will be done with graduate studies by that time."
Mean response: 90.75% (all); 95.11% (males); 87.23% (females) (gender diff. significant at 0.00%)
‡ This is the response to the survey question: "What do you think is the percent chance that you will be active in the FULL-TIME labor force when you are 40 YEARS OLD? (i.e you will be working full-time at the age of 40)
Mean response: 87.91% (all); 92.94% (males); 84.13% (females) (gender diff. significant at 0.01%)

Table A.10. Best Linear Predictor of Expectations of being active in the labor force

	At age of 30[a]			At age of 40		
	All	**Male**	**Female**	**All**	**Male**	**Female**
Major Declared[b]	−3.46*	−2.67*	−5.70*	−4.11*	−4.61*	−5.16
	(1.81)	(1.49)	(3.03)	(2.27)	(2.36)	(3.62)
Cumulative GPA	4.23	1.75	9.36**	8.65***	5.39	16.41***
	(2.64)	(2.11)	(4.67)	(3.32)	(3.33)	(5.58)
SAT_Math[c]	1.19	2.24**	1.18	0.647	0.69	1.49
	(0.88)	(0.92)	(1.37)	(1.11)	(1.45)	(1.64)
SAT_Verbal[d]	0.199	−0.66	0.472	−0.657	−0.98	−1.39
	(0.80)	(0.68)	(1.35)	(1.00)	(1.08)	(1.61)
Female	−6.70***	omitted	omitted	−8.72***	omitted	omitted
	(1.80)			(2.25)		
Black	2.53	2.76	5.06	2.84	5.58	5.07
	(3.90)	(4.41)	(5.74)	(4.90)	(6.98)	(6.85)
Hispanic	13.74**	3.37	18.62**	14.98**	7.40	20.17**
	(5.53)	(6.19)	(8.40)	(6.93)	(9.78)	(10.03)
Asian	6.65**	3.41	6.92	10.34***	6.79*	10.68*
	3.01	(2.31)	(5.35)	(3.77)	(3.65)	(6.39)
Foreign[e]	−2.95	−0.043	−0.55	−9.81	−4.89	−14.56
	(4.95)	(3.69)	(9.34)	(6.21)	(5.84)	(11.16)
Second-Gen Immigrant[f]	−5.29*	−2.19	−8.85*	−8.34**	−7.16*	−11.54*
	(2.88)	(2.32)	(5.04)	(3.61)	(3.68)	(6.02)
Parents' Earnings:[g]						
$75,000-$150,000	1.067	1.76	−0.633	2.32	0.36	3.01
	(2.34)	(1.75)	(4.07)	(2.93)	(2.77)	(4.86)
$150,000-$350,000	4.66*	6.64***	6.12	1.83	1.70	4.61
	(2.43)	(1.88)	(4.12)	(3.05)	(2.97)	(4.92)
$350,000-$500,000	−4.08	3.77	−9.31	−10.36**	−1.37	−14.56*
	(4.06)	(3.37)	(6.40)	(5.10)	(5.33)	(7.65)
> than $500,000	7.15**	7.18**	5.97	8.43**	6.36	10.97*
	(3.21)	(2.73)	(5.20)	(4.02)	(4.31)	(6.21)
Father went to College	−2.43	−3.65	1.21	−5.22	−2.65	−4.88
	(3.67)	(3.35)	(5.90)	(4.61)	(5.30)	(7.04)
Mother went to College	2.04	1.13	0.74	−0.095	−2.09	−1.36
	(2.95)	(2.79)	(4.63)	(3.70)	(4.42)	(5.53)
Mother full-time housewife	−2.00	0.333	−5.97*	−5.02*	1.83	−12.51***
	(2.10)	(1.57)	(3.60)	(2.63)	(2.48)	(4.38)
Parents divorced/ separated	2.69	−2.50	5.75	2.13	−3.76	6.76
	(2.34)	(1.90)	(3.84)	(2.93)	(3.00)	(4.59)
R-squared	0.2793	0.2060	0.2697	0.1840	0.2175	0.1818
No. of Observations	161	69	92	161	69	92

Notes for Table A.10
Parameter estimates correspond to the estimation of a OLS model. Standard errors in parentheses.

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

a Dependent variable is a response 0-100 to: "What do you think is the percent chance that you will be active in the FULL-TIME labor force when you are 30 YEARS OLD?"

b a dummy variable that equals one if the respondent has already declared his/ her major

c (*d*) - SAT Math (Verbal) score; = 1 if SAT Math (Verbal) score is less than 400; =2 if score = 400-499; =3 if score = 500-549; =4 if score = 550-599; =5 if score = 600-649; =6 if score = 650-699; =7 if score =700-749; =8 if score=750-800

e a dummy that equals one if the respondent is an International student

f a dummy that equals one if either of the respondent's parents are foreign-born, and the respondent was born in the US

g The left out income category is Parents' annual earnings is less than $75,000

Table A.11. Correlation Patterns between respondent's and parents' majors (Frequency in cell)

Panel A

Respondent's Major:	*Father's Major* Natural Science	Math & Comp Sc	Social Sci. I	Social Sci. II	Ethics & Values	Area Std	Lit & Arts	Music	Educ	Comm Std	Eng	Journ	N/A[a]	*Total*
Natural Sc.	12	1	1	3	–	–	1	–	–	–	9	–	6	33
Math & Comp.	2	–	1	–	1	–	–	1	–	–	1	1	–	7
Soc. Sci I	10	–	7	9	4	–	1	1	1	–	13	3	12	61
Soc. Sci II	8	5	10	15	4	–	1	–	–	–	19	–	7	69
Eth. & Values	2	–	–	–	1	–	1	–	–	–	3	–	–	7
Area Studies	5	–	4	2	2	–	2	–	–	–	8	–	8	31
Lit. & Fine Arts	4	–	2	3	4	–	3	–	–	–	2	–	9	27
Music Studies	–	–	–	–	–	–	–	–	–	–	1	–	3	4
Education	–	–	–	1	1	–	–	–	–	–	–	–	–	2
Comm. Std.	2	–	–	1	–	–	–	–	–	–	–	–	1	4
Engineering	–	1	–	1	–	–	–	–	–	–	3	–	–	5
Journalism	1	–	–	1	–	–	1	–	–	–	–	–	–	3
Observations	46	7	16	36	17	–	10	2	1	–	59	4	46	253

Panel B

Respondent's Major	*Mother's Major* Natural Science	Math & Comp Sc	Social Sci. I	Social Sci. II	Ethics & Values	Area Std	Lit & Arts	Music	Educ	Comm Std	Eng	Journ	N/A	*Total*
Natural Sc.	13	1	2	5	1	–	3	–	1	–	1	–	6	33
Math & Comp.	3	–		–	–	–	1	1	1	–	1	–	–	7
Soc. Sci I	12	–	10	6	–	–	4	2	7	2	4	1	11	59
Soc. Sci II	11	2	5	12	2	–	8	2	6	1	8	–	14	71
Eth. & Values	2	–	–	–	–	–	3	–	–	–	–	–	1	6
Area Studies	9	1	4	3	–	1	4	1	1	1	2	–	5	32
Lit. & Fine Arts	6	–	–	3	–	–	6	–	6	–	–	–	6	27
Music Studies	–	–	–	–	–	–	1	–	–	–	–	–	3	4
Education	–	–	2	–	–	–	–	–	1	–	–	–	–	3
Comm. Std.	1	–	–	–	1	–	–	–	–	–	–	–	1	3
Engineering	–	–	–	1	–	–	–	–	1	1	1	–	1	5
Journalism	1	–	–	–	–	–	–	–	–	1	–	–	1	3
Observations	58	4	23	30	4	1	30	6	24	6	17	1	49	253

a Includes the case where the parent's major is not known, or does not exist.

APPENDIX B

Appendix 2 for Chapter 2

B.1. Tables for Choice Model Estimation

Table B.1. Single Major Choice- Estimation of Homogeneous Preferences

	Using Stated Choice		*Using stated Preference*	
	(1)	(2)	(3)	(4)
Δu_1 for graduating within 4 years	6.84*** (1.78)	1.65 (2.93)	−0.447 (0.868)	−1.54* (0.80)
Δu_1 for graduating within 4 years × female	–	54.27*** (6.63)	–	3.15** (1.37)
Δu_2 for graduating with a GPA of at least 3.5	−3.83*** (1.11)	−1.95 (1.94)	0.903* (0.520)	1.13* (0.67)
Δu_2 for graduating with a GPA of at least 3.5 × female	–	−8.44** (4.03)	–	0.048 (1.12)
Δu_3 for enjoying the coursework	13.11*** (2.47)	9.93** (4.36)	2.69*** (0.45)	2.06*** (0.70)
Δu_3 for enjoying the coursework × female	–	11.36 (8.39)	–	1.43 (0.946)
γ_1 for hours/week spent on coursework	−0.058*** (0.017)	−0.057** (0.028)	0.012 (0.011)	0.0064 (0.0135)
γ_1 for hours/week spent on coursework × female	–	−0.045 (0.071)	–	0.0189 (0.021)
Δu_4 for approval of parents and family	3.71*** (1.16)	1.74 (3.14)	1.37** (0.56)	0.98 (0.75)
Δu_4 for approval of parents and family × female	–	1.71 (3.98)	–	1.03 (1.13)
Δu_5 for finding a job upon graduation	2.27* (1.20)	4.01* (2.17)	−0.076 (0.512)	0.279 (0.829)
Δu_5 for finding a job upon graduation × female	–	0.74 (4.15)	–	−0.863 (1.04)
Δu_6 for enjoying work at the available jobs	6.65*** (2.05)	1.80 (3.21)	1.59*** (0.384)	0.468 (0.526)
Δu_6 for enjoying work at the available jobs × female	–	18.86*** (7.01)	–	1.80** (0.817)
Δu_7 for reconciling family and work at available jobs	−1.93* (1.11)	−1.31 (2.77)	0.241 (0.539)	0.258 (0.671)
Δu_7 for reconciling family and work × female	–	−2.36 (4.66)	–	0.181 (0.946)
γ_2 for hours/week spent at work	−0.0066 (0.0166)	0.0282 (0.038)	−0.0080 (0.0099)	−0.015 (0.015)
γ_2 for hours/week spent at work × female	–	−0.073 (0.082)	–	0.024 (0.018)
γ_3 for the social status of the available jobs[a]	3.27*** (1.12)	4.01* (2.28)	1.09*** (0.32)	2.14*** (0.53)
γ_3 for the social status of the available jobs × female	–	−0.59 (4.08)	–	−1.696** (0.662)
γ_4 for expected Income at the age of 30	−5.25e − 7 (4.25e − 6)	9.43e − 6 (7.91e − 6)	6.43e − 7 (1.02e − 6)	1.13e − 6 (2.43e − 6)
γ_4 for expected Income at the age of 30 × female	–	−19.1e − 6 (21.8e − 6)	–	−4.40e − 7 (2.53e − 6)
Log-Likelihood	−56.58	−40.77	−733.52	−703.255
No. of Observations	83	83	83	83

* significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses
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) are probabilities between 0 and 1

Table B.2. Decomposition Analysis

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

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 B.3. Thought Experiments

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

[a] each cell corresponds to the percent change in the probability of majoring in that category after the intervention relative to the baseline case.

Table B.4. Estimation of heterogeneous preferences using Stated Preference

	Entire Sample	Males	Females
	(1)	(2)	(3)
Δu_1 for graduating within 4 years	-0.545	-0.958	1.20
	(0.791)	(0.911)	(1.21)
Δu_2 for graduating with a GPA of at least 3.5	0.752	0.751	1.01
	(0.575)	(0.721)	(1.01)
Δu_3 for enjoying the coursework	2.92^{***}	2.49^{***}	3.57^{***}
	(0.466)	(0.754)	(0.658)
γ_1 for hours/week spent on coursework[a]	0.0152	0.0098	0.0232
	(0.011)	(0.014)	(0.016)
Δu_4 for parents approv $\times$ parents'_supp[d] $\times$ (1-Foreign[e])	0.340^{**}	0.578^{***}	0.262
	(0.150)	(0.217)	(0.194)
$\widetilde{\Delta u_4}$ for parents approval $\times$ parents'_support $\times$ Foreign	0.0439	-0.147	0.601^{**}
	(0.159)	(0.205)	(0.246)
Δu_5 for finding a job upon graduation	0.205	0.680	-0.536
	(0.494)	(0.759)	(0.637)
Δu_6 for enjoying work at the available jobs	1.51^{***}	0.319	2.24^{***}
	(0.414)	(0.611)	(0.678)
Δu_7 for reconciling family and work at available jobs	0.246	0.700	0.547
	(0.579)	(0.747)	(0.847)
$\widetilde{\Delta u_7}$ for reconciling family & work $\times$ divorced[f]	-0.357	0.494	-0.613
	(0.864)	(1.26)	(1.32)
γ_2 for hours/week spent at work[c]	-0.0097	-0.0044	0.0045
	(0.0100)	(0.016)	(0.012)
γ_3 for social status of the available jobs[b] $\times$ (1-Foreign)	0.310	1.30^{*}	0.297
	(0.432)	(0.76)	(0.546)
$\widetilde{\gamma_3}$ for social status of jobs $\times$ Foreign	2.28^{***}	3.27^{***}	0.817
	(0.550)	(0.93)	(0.580)
γ_4^{HI} for exp. Inc at 30 $\times$ (1- low_inc[g]) $\times$ (1-Foreign)	$2.66e-06$	$3.08e-06$	$17.5e-06$
	$(2.75e-06)$	$(2.80e-06)$	$(12.5e-06)$
$\widetilde{\gamma_4^{HI}}$ for exp Inc at 30 $\times$ (1-low_income) $\times$ Foreign	$-8.16e-07$	$-11.1e-06$	$7.13e-07$
	$(2.33e-06)$	$(8.07e-06)$	$(7.28e-06)$
γ_4^{LI} for exp. Income at 30 $\times$ low_inc $\times$ (1-Foreign)	$1.06e-07$	$-3.89e-06$	$1.02e-06$
	$(3.39e-06)$	$(3.54e-06)$	$(2.58e-06)$
$\widetilde{\gamma_4^{LI}}$ for expected Income at 30 $\times$ low_inc $\times$ Foreign	$6.64e-06$	$11.3e-06^{**}$	$1.40e-06$
	$(4.55e-06)$	$(5.42e-06)$	$(5.78e-06)$
Log-Likelihood	-726.19	-287.61	-401.68
No. of groups	83	33	50

Notes for Table B.4

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

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

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

c - social status is on a scale of 1-8 (8 being the highest social status); normalized to be between 0.1-0.8 all other variables (except income) are probabilities between 0 and 1

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

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

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

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

Table B.5. Decomposition Analysis

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

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

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

Table B.6. Best Linear Predictor of Expectation of Parent's Approval

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

Note for Table B.6

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

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

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

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

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

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

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

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

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

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

Table B.7. Mixed Logit Model Estimation

		Without Demographics			With Demographics		
			Mean, Dev Log-Normal[a]			Mean, Dev Log-Normal	
		Est. (Error	Mean	Dev.	Est. (Error)	Mean	Dev.
		(1*a*)	(1*b*)	(1*c*)	(2*a*)	(2*b*)	(2*c*)
Fixed Coefficients							
Δu_2 for graduating within 4 years	Coeff	−0.12 (0.69)			0.042 (0.73)		
$\underset{\sim}{\gamma_1}$ hours/week on coursework	Coeff	0.021* (0.01)			0.024* (0.01)		
$\widetilde{\Delta u_4}$ approval × parent_sup × For.	Coeff	–			−0.51** (0.21)		
Δu_5 finding a job upon graduation	Coeff	−0.13(0.46)			0.067 (0.47)		
$\underset{\sim}{\Delta u_7}$ reconcile family & work at job	Coeff	0.074 (0.49)			0.013 (0.54)		
$\widetilde{\Delta u_7}$ reconcile family & work× Div	Coeff	–			0.391 (1.01)		
γ_2 for hours/week spent at work	Coeff	−0.01 (0.01)			−0.65 (0.77)		
γ_4 - Expected Income at 30	Coeff	5e-7 (1.2e-6)			−7e-7 (1.5e-6)		
Random Coefficients							
Δu_2 graduating with GPA ≥ 3.5	$\overline{\Delta u}$	−0.55 (1.16)	2.31	8.14	−1.01 (1.24)	2.88	16.88
	σ	1.69* (0.87)			2.07*** (0.67)		
Δu_3 for enjoying the coursework	$\overline{\Delta u}$	1.09*** (0.20)	4.01	3.62	1.18*** (0.19)	4.23	3.52
	σ	0.79*** (0.19)			0.74*** (0.18)		
Δu_4 for approval of parents	$\overline{\Delta u}$	0.179 (0.485)	2.11	3.11	–	–	–
	σ	1.075*** (0.37)					
Δu_4 approval × Parents'_supp	$\overline{\Delta u}$	–	–	–	−0.662* (0.383)	0.725	0.730
	σ				0.835*** (0.287)		
Δu_6 for enjoying work at jobs	$\overline{\Delta u}$	0.404 (0.388)	2.28	2.64	0.441 (0.387)	2.25	2.38
	σ	0.912*** (0.396)			0.857** (0.408)		
γ_3 for the social status of jobs	$\overline{\Delta u}$	−0.493 (0.67)	2.19	9.18	−1.66 (2.17)	2.25	31.35
	σ	1.59*** (0.545)			2.19** (1.04)		
$\widetilde{\gamma_3}$ social status × Foreign	Mean	–	–	–	2.53*** (0.84)		
	St. Dev				1.98** (0.88)		
$\widetilde{\gamma_4}$ Income at 30 × Low_Inc	Mean	–	–	–	1.57e-6 (2.9e-6)		
	St. Dev.				2.84e-6 (5.1e-6)		
Log-Likelihood		−698.00			−690.84		
No. of Groups		83			83		
No. of Observations		2957			2957		

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

Table B.8. Decomposition Analysis for double major respondents using stated preference data

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

Table B.9. Distribution of Double Majors

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

Table B.10. Double Major Choice Model - Estimation Using Choice Data

	All		**Males**		**Females**	
	Estimate	Error	Estimate	Error	Estimate	Error
Variables	(1*a*)	(1*b*)	(2*a*)	(2*b*)	(3*a*)	(3*b*)
Δu_{11} for graduating within 4 years	−1.81	(2.73)	−5.92	(3.89)	−1.59	(5.00)
Δu_{12} for maximum of graduating in 4 years	23.70***	(3.37)	19.77***	(6.53)	24.17***	(5.66)
Δu_{2} for graduating with a GPA of $\geq$3.5	3.53***	(1.19)	2.27	(1.98)	5.54***	(1.71)
Δu_{31} for enjoying the coursework	11.54***	(2.74)	10.54**	(4.75)	12.15**	(4.60)
Δu_{32} for maximum of enjoying coursework	3.41	(3.02)	24.85***	(6.42)	−1.38	(4.24)
γ_{11} for hours/week spent on coursework	0.155***	(0.043)	0.190	(0.134)	0.157***	(0.061)
γ_{12} for min. of hours/week on coursework	−0.176***	(0.047)	−0.101	(0.151)	−0.217***	(0.059)
Δu_{41} for approval of parents and family	9.17***	(2.03)	6.70	(4.44)	10.00***	(3.26)
Δu_{42} for maximum of approval of parents	1.58	(2.43)	10.02*	(6.14)	0.52	(3.64)
Δu_{51} for finding a job upon graduation	−3.31*	(1.78)	−1.84	(3.12)	−4.36	(1.40)
Δu_{52} for maximum of finding a job	5.67***	(2.01)	5.46	(3.70)	6.48**	(2.72)
Δu_{61} for enjoying work at the jobs	4.51***	(1.13)	6.01**	(2.81)	4.33***	(1.40)
Δu_{71} for reconciling family & work at jobs	1.85**	(0.91)	0.151	(2.22)	2.23**	(1.11)
γ_{21} for hours/week spent at work	0.047***	(0.015)	0.021	(0.034)	0.033	(0.021)
γ_{31} for the social status of the jobs	−0.644	(0.82)	0.935	(2.00)	−1.08	(1.11)
γ_{41} for expected income at the age of 30	1.1e−6	(2.8e−6)	1.3e−6	(7.1e−6)	5.7e−7	(3.7e−6)
Log Likelihood	−154.87		−59.66		−86..43	
Number of Respondents	78		36		42	

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

Table B.11. Contributions of various outcomes

	All	**Males**	**Females**
	Estimates	Estimates	Estimates
Variables	(1)	(2)	(3)
$\Delta u_{11} + \Delta u_{12}$ (graduating within 4 years)	21.88***	13.84**	22.58***
Δu_2 (graduating with a GPA of at least 3.5)	3.53***	2.27	5.54***
$\Delta u_{31} + \Delta u_{32}$ (enjoying the coursework)	14.95***	35.39***	10.77***
$\gamma_{11} + \gamma_{12}$ (hours/week spent on coursework)	−0.0200	0.089**	−0.0597*
$\Delta u_{41} + \Delta u_{42}$ (approval of parents and family)	10.75***	16.73***	10.52***
$\Delta u_{51} + \Delta u_{52}$ (finding a job upon graduation)	2.36**	3.62*	2.12*
$\Delta u_{61} + \Delta u_{62}$ (enjoying work at the available jobs)	4.51***	6.01**	4.34***
$\Delta u_{71} + \Delta u_{72}$ (reconciling family and work at the jobs)	1.85**	0.151	2.23**
$\gamma_{21} + \gamma_{22}$ (hours/week spent at work)	0.047***	0.021	0.0326
$\gamma_{31} + \gamma_{32}$ (social status of the available jobs)	−0.644	0.93	−1.08
$\gamma_{41} + \gamma_{42}$ (expected income at the age of 30)	1.09e−06	1.33e−06	5.72e−07

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

Table B.12. Double Major Choice Model with Error Components

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

† Includes Error Components for each major category (estimates not shown)
* significant at 10%; ** significant at 5%; *** significant at 1%

Table B.13. 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: The contribution in % of the relevant variable to the gap

Table B.14. Simulations of the 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	0.0608	0.0308	0.0534
		13.92%[b]	−1.06%	48.74%	11.18%
Natural Sciences	0.0550	0.0445	0.0529	0.0229	0.0406
		18.98%	3.88%	58.29%	26.48%
Math & Computer Sci.	0.0191	0.0135	0.0184	0.0074	0.0083
		29.07%	3.45%	61.41%	56.38%
Social Sciences I	−0.0412	−0.0524	−0.0474	−0.0643	−0.0613
		−27.28%	−15.32%	−56.25%	−48.84%
Social Sciences II	0.0907	0.0737	0.0881	0.0272	0.0608
		18.68%	2.88%	69.92%	32.92%
Ethics & Values	−0.0189	−0.0266	−0.0219	−0.0419	−0.0381
		−40.77%	−15.87%	−122.03%	−101.9%
Area Studies	−0.0624	−0.0634	−0.0655	−0.0563	−0.0721
		−1.69%	−4.96%	9.87%	−15.48%
Lit. & Fine Arts	−0.0843	−0.0863	−0.0888	−0.0545	−0.0777
		-2.35%	−5.35%	35.34%	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

APPENDIX C

Appendix for Chapter 3

C.1. Experimental Instructions

Welcome to the Experiment. Thanks you for your participation.

Instructions

You are about to participate in an experimental study of decision making. The experiment will last about half an hour. In the experiment, you will be assigned randomly to a group consisting of 4-6 persons (you will get to know the actual number when the experiment starts). Your group will remain the same for the entire experiment. As you notice, there are more people in the room than the size of the group, so you cannot know who are the other members of your group. A person sitting next to you may or may not be in the same group as you. Please try not to look at other people's screens during the experiment. You will be paid $4 as a show-up fee. In addition, you may earn money during the experiment.

The experiment will consist of 6 rounds. At the end of the experiment, one round will be picked at random to determine your earnings from the experiment. Therefore, you should treat each round as a real round. All the money will be paid in cash at the end of the experiment.

You will not be allowed to talk or communicate with other participants. If you have a question, please raise your hand and I will come to you. Are there any questions about what has been said up till now?

The next section describes the basic idea of the experiment.

The Basic Procedure

Recently the world, and in particular the US, has been struck by several natural disasters. Currently, there are hundreds of thousands of hurricane survivors (US), famine survivors (Africa), and flood-displaced individuals (Asia) out there who are in dire need of your help. Over 1 billion people—1 in 6 people around the world—live in extreme poverty, defined as living on less than $1 a day. More than 800 million go hungry each day. This experiment provides you with the opportunity to help a cause that might be dear to your heart.

You will get $10 every round, and will have the opportunity to donate part of this amount to the Red Cross (you will get to choose the specific effort to which your donation will go). The decision procedure will be as follows: you will have two options:

1 Take the $10, and do not donate any amount to the Red Cross. In this case, your earnings for that period will be $10.

2 Donate part of your $10 to the Red Cross. If you decide to donate x (x can be any number between 0.5 and 10 in multiples of $0.5), then one of two things can happen:

   – With probability 60%, the donation goes through: Red Cross will get x, and you get $\$10 - x$.

   – With probability 40%, the donation does not go through, i.e. Red Cross does not get anything, and you keep your $10.

If you decide to donate, you will be given the option to choose the specific Red Cross effort to which you want to contribute.

Note there are two ways in which Red Cross does not get anything: either you decide not to donate anything; or you decide to donate, but the donation does not go through.

The instructions for each student in your group will be the same for each round.

This procedure will be repeated for 6 rounds. Each round might have some modification about which you will be informed before you make the decision.

At the end of the experiment, one round will be picked at random. You will be paid in cash for your earnings of that round (as well as the show-up fee), and Red Cross will receive the donation for that round (you will be given a receipt of the donation made to Red Cross).

The next section describes some of the efforts undertaken by the Red Cross to which you may direct your donation.

About the Red Cross

The American Red Cross is comprised of hundreds of local Red Cross chapters and blood services regions that provide a variety of programs and services in cities, towns and neighborhoods across the country and around the world. You could direct your gift to any of the Red Cross efforts listed below:

(1) NATIONAL DISASTER RELIEF FUND: The American Red Cross responds to approximately 75,000 disasters a year, including Tropical Storms Dean and Erin, floods, house fires, storms, tornadoes, hurricanes and other disasters, providing aid to more disaster victims nationwide. This means that every eight minutes a disaster strikes and a family turns to the Red Cross for help. The Red Cross stands ready to turn the compassion of our donors into action. With your support of the Disaster Relief Fund, a fund that requires constant replenishment, the Red Cross can be there for the victims of the recent storms.

(2) WHERE OUR NEED IS GREATEST: The American Red Cross is where people mobilize to help their neighbors—down the street, across the country and around the world—in emergencies. The American Red Cross, a humanitarian organization led by

volunteers, guided by its Congressional Charter and the Fundamental Principles of the International Red Cross Movement, provides relief to victims of disasters and helps people prevent, prepare for, and respond to emergencies. You can help ensure that the Red Cross can continue to provide these lifesaving services and has the resources, talent and ability to continue to deliver them by making a donation to support all of its core services today.

(3) INTERNATIONAL RESPONSE FUND: You can help those affected by countless crises around the world each year by making a financial gift to the American Red Cross International Response Fund, which will provide immediate relief and long-term support through supplies, technical assistance and other support to help those in need. If you wish to designate your donation to a specific disaster please do so at the time of your donation.

(4) YOUR LOCAL RED CROSS CHAPTER: Your local Red Cross chapter is committed to meeting the humanitarian needs of the people in your area, be it in disaster preparedness, disaster response, first aid and CPR training, or disease prevention. You can help support your local chapter programs and services by a gift to Your Local Red Cross Chapter. The gift will be sent to your local area chapter based on zip code.

(5) MILITARY SERVICES: The American Red Cross is a lifeline for deployed military members, allowing them to communicate to loved ones back home during emergencies. You can help the Red Cross keep military families connected with a gift to Red Cross Armed Forces Emergency Services.

(6) MEASLES INITIATIVE: The Measles Initiative is a partnership committed to reducing measles deaths globally. Launched in 2001, the Measles Initiative—led by the

American Red Cross, the United Nations Foundation, the U.S. Centers for Disease Control and Prevention, UNICEF and the World Health Organization—provides technical and financial support to governments and communities on vaccination campaigns in all regions of the world. To date, the Initiative has supported the vaccination of more than 372 million children helping to reduce measles deaths by more than 60% globally (compared to 1999). To learn more, visit http://www.measlesinitiative.org/.

(7) BLOOD SERVICES CAMPAIGN: Your gift to the Blood Services Campaign of the American Red Cross supports our commitment to the nation's blood supply. Through this 10-year undertaking we will update and reconfigure our blood manufacturing facilities across the nation, to better serve the health needs of patients nationwide.

The instructions for each round are available upon request.

C.2. Marlowe-Crowne 2(10) Social Desirability Scale

This is taken from Mandell.

Respondents were required to answer True or False to the following set of 10 questions:

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you personally.

1. I never hesitate to go out of my way to help someone in trouble. (T)
2. I have never intensely disliked anyone. (T)
3. There have been times when I was quite jealous of the good fortune of others. (F)
4. I would never think of letting someone else be punished for my wrong doings. (T)
5. I sometimes feel resentful when I don't get my way. (F)
6. There have been times when I felt like rebelling against people in authority even

though I knew they were right. (F)

7. I am always courteous, even to people who are disagreeable. (T)

8. When I don't know something I don't at all mind admitting it. (T)

9. I can remember "playing sick" to get out of something. (F)

10. I am sometimes irritated by people who ask favors of me. (F)

The scoring algorithm is as follows:

For each answer the respondent provides that matches the response given above (i.e.,T=T or F=F) assign a value of 1. For each discordant response (i.e., the respondent

provides a T in place of an F or an F in place of a T) assign a value of 0. Total score can range from 10 (when all responses "match") to 0 (when no responses "match"). A higher score implies higher social desirability.

C.3. Debriefing

Below I report selective responses of subjects to the question: "What was your strategy during the course of the experiment? Especially explain the pattern of your donations during the experiment."

- I had originally decided on a set amount to donate, and I saw no reason to change it.
- I did not want to donate $0.0 to the Red Cross because it is a great organization. I wanted to donate something and $1 was 6.6% of my maximum payoff, which I thought was fair. I want to give more than 5% and at the same
- Initially, I reckoned that everyone would donate $10 since it didn't cost them anything. However, after seeing that the average for that round was a measly $3, I lowered my donation to $5.

- I guessed that the more information that each person received about the other participants the more they would choose to donate.
- Earn the most money possible.
- I figured that the $15 would be of more use to me right now as a broke college student than it would be to the Red Cross, so I did not donate at all.
- Try to maximize my earnings.
- The strategy was to maximize payoff.
- It was essentially random. I felt the need to donate more if I knew people could find out whether or not I TRIED to donate
- I donated the same amount each time because I knew how much I wanted to donate and none of the conditions changed that.
- I originally was donating out of generosity but as I saw that others were not doing so, I no longer felt that it was a group effort to donate so I did not in round 4. However, seeing at least some donation after that, I decided to donate
- At first, I thought that 2 dollars was an efficient amount, but after seeing that most people donated around 5 dollars, I felt selfish and increased it a little, especially after seeing someone donate 10 dollars.
- When people began to know who I was, I donated some money to the red cross. Then it turned out that it didn't matter anyway so I donated nothing.
- I wanted to keep the entire 15, and don't like to gamble, so I consistently did not donate, even with the pressure of people knowing. Since I only know one other person in this room, I'm not too worried about being judged for it.

- I wasn't planning on donating until my identity and location became known to other members of my team
- My strategy was donate every time an amount I was willing to lose. On estimating the average given to the Red Cross, I figured that \$2 was safe bet because it was good average based on my assumption of what people would donate.
- I think donating half the money is fair, so I did that. However, at the beginning of round 2 I saw the average was \$6, so I decided I should be more generous like my peers. Then, when they started donating less, I went back to my original mind.
- I kept my donations more or less consistent from round to round but I was a poor judge of the average amount donated by my group. Even though group members knew my identity, I wasn't especially influenced because I did not know them.
- To maintain a consistent donation pattern throughout the experiment.
- I donated a little at the beginning. However, when people could see my identity, I donated slightly more.
- I wanted to be generous but still keep some for myself, so I originally decided to walk away with 10. However, seeing the average donation made me reconsider my donation, and I lowered it to about the average.

C.4. Mathematical Appendix

Proposition 1: Stated in the body of the paper.

Proof. Agent i's maximization problem is:

$$\max_{x\in[0,1]} \mathbf{1}_{SI} * u(1-x_i) + \mathbf{1}_A * u(1-x_i, \sum_{j=1}^{N_i} x_j) + \mathbf{1}_{WG} * u(1-x_i, x_i) + t_i G(\frac{|x_i - s_{-i}|}{\sigma_{-i}}) \tag{C.1}$$

Case 1: $1 > x^{**} > s$:The FOC is:

$$\underbrace{\mathbf{1}_A * u_2 + \mathbf{1}_{WG} * u_2}_{+\,ve} + \underbrace{\frac{t}{\sigma}G^{'}(\frac{|x-s|}{\sigma}) - u_1}_{-ve} = 0 \tag{C.2}$$

Since this FOC is never satisfied for a self-interested agent, there is no $x^{**} > s$. For an agent motivated by altruism or warm glow, this FOC is satisfied.

Case 2: $0 < x^{**} < s$: The FOC is:

$$\underbrace{-u_1}_{-ve} + \underbrace{\mathbf{1}_A * u_2 + \mathbf{1}_{WG} * u_2 - \frac{t}{\sigma}G^{'}(\frac{|x-s|}{\sigma})}_{+ve} = 0 \tag{C.3}$$

This FOC can be satisfied by an $0 < x^{**} < s$. This is the case of partial conformity

Case 3: $x^{**} = 0$. The FOC is:

$$\underbrace{-u_1}_{-ve} + \underbrace{\mathbf{1}_A * u_2 + \mathbf{1}_{WG} * u_2 - \frac{t}{\sigma}G^{'}(\frac{|x-s|}{\sigma})}_{+ve} < 0 \tag{C.4}$$

which can be satisfied.

Case 4: $x^{**} = 1$. The FOC is:

$$\underbrace{\mathbf{1}_A * u_2 + \mathbf{1}_{WG} * u_2}_{+\,ve} + \underbrace{\frac{t}{\sigma}G^{'}(\frac{|x-s|}{\sigma}) - u_1}_{-ve} > 0 \tag{C.5}$$

which is never satisfied for a self-interested agent, but is satisfied for an altruistic or warm glow motivated agent.

Case 5: $x^{**} = s$.Recall that $G(0) = 0$. The FOC is simply:

$$\underbrace{-u_1}_{-\,ve} + \underbrace{\mathbf{1}_A * u_2 + \mathbf{1}_{WG} * u_2}_{+ve} = 0 \tag{C.6}$$

which is never satisfied for a self-interested agent, but is satisfied for an altruistic or warm glow motivated agent.

Thus, $x^{**}(t) \in [0, s)$ for a self-interested agent, while $x^{**}(t) \in [0, 1]$ for an agent motivated by altruism or warm glow. □

Proposition 2: Stated in the body of the paper.

Proof. Proof of (a):

I show the proof for the case of a self-interested agent. For the other two types, one only needs to replace the term $u_{11}(1-x)$ with $u_{11}(1-x,\zeta) + u_{22}(1-x,\zeta)$ where ζ is the relevant argument. The proof still goes through because the utility is concave in both arguments.

(1) Differentiating equation C.3 implicitly yields:

$$\frac{dx}{dt} = \frac{\frac{1}{\sigma}G'(\frac{|x-s|}{\sigma})}{u_{11}(1-x) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.7}$$

under the concavity assumptions on $u(.)$, and $G(.)$, and the assumption that $u_{12} < 0$, both the numerator and the denominator are strictly negative, and thus $\frac{dx}{dt} > 0$. Moreover, note that if $x = 0$ is optimal for some t, then from equation C.4 it is also optimal for smaller t.

(2) Implicitly differentiating equation C.3 with respect to s:

$$\frac{dx}{ds} = \frac{\frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})}{u_{11}(1-x) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.8}$$

since both the numerator and denominator are strictly negative, $\frac{dx}{ds} > 0$

(3) Implicitly differentiating equation C.3 with respect to σ:

$$\frac{dx}{d\sigma} = -\frac{\frac{t}{\sigma^2}G'(\frac{|x-s|}{\sigma}) + \frac{t|x-s|}{\sigma^3}G''(\frac{|x-s|}{\sigma})}{u_{11}(1-x) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.9}$$

the numerator and denominator are both negative. With the negative sign in front of them, $\frac{dx}{d\sigma} < 0$.

Proof of (b):

(1) Differentiating equation C.2 implicitly yields:

$$\frac{dx}{dt} = \frac{-\frac{1}{\sigma}G'(\frac{|x-s|}{\sigma})}{u_{11}(1-x,\zeta) + u_{22}(1-x,\zeta) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.10}$$

since the numerator is positive, and the denominator is strictly negative, $\frac{dx}{ds} < 0$. Moreover, if $x = 1$ is optimal for some t, then from equation C.5 it is also optimal for larger t.

(2) Implicitly differentiating equation C.2 with respect to s:

$$\frac{dx}{ds} = \frac{\frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})}{u_{11}(1-x,\zeta) + u_{22}(1-x,\zeta) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.11}$$

the numerator and denominator are both strictly negative, and so $\frac{dx}{ds} > 0$.

(3) Implicitly differentiating equation C.2 with respect to σ:

$$\frac{dx}{d\sigma} = \frac{\frac{t}{\sigma^2}G'(\frac{|x-s|}{\sigma}) + \frac{t|x-s|}{\sigma^3}G'(\frac{|x-s|}{\sigma})}{u_{11}(1-x,\zeta) + u_{22}(1-x,\zeta) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})} \tag{C.12}$$

the numerator and denominator are both strictly negative, and so $\frac{dx}{d\sigma} > 0$. □

Claim 4: Stated in the body of the paper

Proof. In order to prove this claim, I need to show that:

(1) For the case where $\widetilde{x(t)} < s$: the full information contributions weakly increase in t at a faster rate than they would in the limited information if perception and consumption are substitutes (i.e. $u_{12} < 0$ for a self-interested agent; $u_{13} < 0$ and $u_{23} > 0$ for the other two types). I show the proof for the case of a self-interested agent. In a full information case, equation C.7 now becomes

$$\frac{dx}{dt} = \frac{u_{12}(1-x,t) + \frac{t}{\sigma}G'(\frac{|x-s|}{\sigma})}{u_{11}(1-x,t) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})}$$

under the assumption that $u_{12} < 0$, then both the numerator and denominator are strictly negative. Thus $\frac{dx}{dt}$ is positive and strictly larger than the slope in equation C.7.

(2) For the case where $\widetilde{x(t)} > s$, the contribution in the full information case decreases in t at a slower rate than it would in the limited information case if it is assumed that consumption and perception are substitutes ($u_{13} < 0$), and that contribution and perception are compliments ($u_{23} > 0$). In order to show this, note that in the full information

case, equation C.10 is:

$$\frac{dx}{dt} = \frac{u_{13}(1-x,\varsigma,t) - u_{23}(1-x,\varsigma,t) - \frac{1}{\sigma}G'(\frac{|x-s|}{\sigma})}{u_{11}(1-x,\zeta) + u_{22}(1-x,\zeta) + \frac{t}{\sigma^2}G''(\frac{|x-s|}{\sigma})}$$

The denominator is strictly negative. Under the assumption that $u_{13} < 0$ and $u_{23} > 0$, the numerator may be positive or negative. However, $\frac{dx}{dt}$ is strictly greater than the slope in equation C.10.

These two claims collectively imply that the contribution distribution in the full information case will first-order stochastically dominate the cumulative distribution function of contributions in the limited information case. □

Claim 5: Stated in the body of the paper

Proof. From equation C.4,

$$t < \frac{-u_1(1,t)}{\frac{1}{\sigma}G'(\frac{s}{\sigma})}$$

So equation C.4 is satisfied for $t^* = \min\{\frac{-u_1(1,t)}{\frac{1}{\sigma}G'(\frac{s}{\sigma})}, \bar{t}\}$. Similarly from equation C.5;

$$t > \frac{-u_1(1,t) + \mathbf{1}_{WG} * u_2 + \mathbf{1}_A * u_2}{\frac{1}{\sigma}G'(\frac{s}{\sigma})}$$

So equation C.5 is satisfied for $\widehat{t}$ =max$\{\frac{-u_1(1,t)+\mathbf{1}_{WG}*u_2+\mathbf{1}_A*u_2}{\frac{1}{\sigma}G'(\frac{s}{\sigma})}, 0\}$. □

Proposition 5: Stated in the body of the paper.

Proof. The proof follows from Bernheim (1994), and requires one to write down the individual rationality constraints to avoid mutual imitation. Consider two types, t and $t'$with $t > t'$. Suppose type t chooses x earning a perception of p,while type t' chooses x'

earning a perception of p'. The IC constraints are:

$$u(1-x,x,p)+tG(\frac{|x-s|}{\sigma})\geq u(1-x',x',p')+tG(\frac{|x'-s|}{\sigma}) \tag{C.13}$$

and

$$u(1-x',x',p')+t'G(\frac{|x'-s|}{\sigma})\geq u(1-x,x,p)+t'G(\frac{|x-s|}{\sigma}) \tag{C.14}$$

which implies

$$(t-t')[G(\frac{|x-s|}{\sigma})-G(\frac{|x'-s|}{\sigma})]\geq 0 \tag{C.15}$$

Since $t>t'$, it must be that $G(\frac{|x-s|}{\sigma})\geq G(\frac{|x'-s|}{\sigma})$, which implies that x is closer to s than x' is to s, i.e. x is a more conforming choice. $\square$

C.5. Figures and Tables

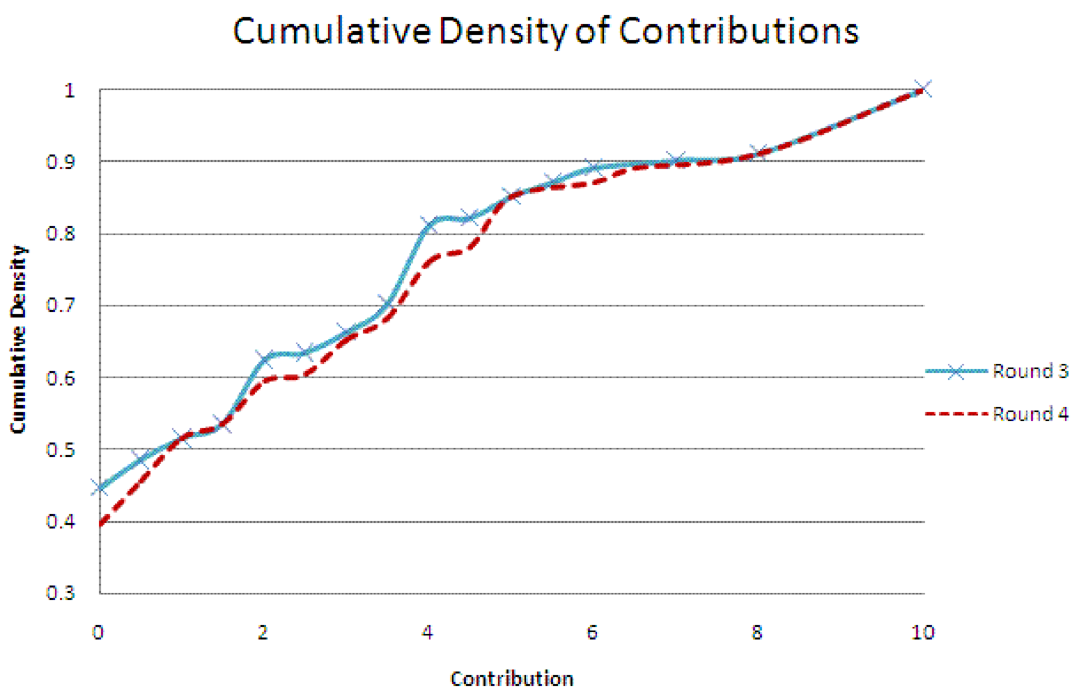

Figure C.1. Cumulative Density of Contributions

Table C.1. Donation Statistics

	Number of subjects who donate	Number of subjects who donate all endowment	Avg. amount contributed	Avg. amount contributed by subjects who donate
Round 1	56	9	$2.33	$4.21
Round 2	57	7	$2.15	$3.81
Round 3	56	9	$2.41	$4.33
Round 4	61	9	$2.56	$4.25
Round 5	54	9	$2.39	$4.47
Round 6	55	8	$2.31	$4.24

Table C.2. Wilcoxon Rank Sum Test

	Test Statistic	p-value
Round 2 - Round 1	0.365	0.715
Round 3 - Round 2	1.231	0.218
Round 4 - Round 3	1.970	0.048
Round 5 - Round 4	-1.351	0.177
Round 6 - Round 5	-0.863	0.388

Table C.3. Contribution response to group choice in Round 2

	OLS	Tobit	OLS	Tobit
	(1)	(2)	(3)	(4)
α (constant)	-0.67^{**}	-0.70^{**}	0.15	0.16
	(0.32)	(0.34)	(0.24)	(0.25)
β (soc. comparison)	0.31^{**}	0.32^{***}	–	–
	(0.11)	(0.12)		
β_1(soc. comparison if donated LESS than group avg. in round 1)	–	–	0.024	0.021
			(0.07)	(0.07)
β_2 (soc. comparison if donated MORE than group avg. in round 1)	–	–	-0.56^{**}	-0.59^{**}
			(0.24)	(0.26)

NOTE: robust standard errors in parentheses; * sig. at 1%, ** sig. at 5%; *** sig. at 10%

Table C.4. Contribution response to group choice in Round 3

	OLS	Tobit	OLS	Tobit
	(1)	(2)	(3)	(4)
α (constant)	0.016	0.0187	−0.385	−0.389
	(0.191)	(0.192)	(0.329)	(0.313)
β (soc. comparison)	0.157*	0.159*	–	–
	(0.081)	(0.084)		
β_1 (soc. comparison effect if donated LESS than group avg. in round 2)	–	–	0.297**	0.302**
			(0.134)	(0.137)
β_2 (soc. comparison effect if donated MORE than group avg. in round 2)	–	–	0.025	0.0259
			(0.058)	(0.057)

Table C.5. Contribution response to group choice in Rd. 3

	Tobit	Tobit
	(1)	(2)
α (constant)	0.005	-0.513
	(0.202)	(0.318)
$\beta \times$ 1[unchanged in round 2][a]	0.141	–
	(0.091)	
$\beta \times$ (1-1[unchanged in round 2])	0.221*	–
	(0.131)	
$\beta_1\times$ 1[unchanged in round 2]	–	0.297**
		(0.143)
$\beta_1\times$ (1-1[unchanged in round 2])	–	0.426***
		(0.159)
$\beta_2\times$ 1[unchanged in round 2]	–	0.005
		(0.062)
$\beta_2\times$ (1-1[unchanged in round 2])	–	0.334
		(0.309)

$a:$ 1[unchanged in round 2]=1 if subject's round 2 contribution is same as in round 1

Table C.6. Contributions and the Group Choice

Dependent Variable: Contribution (x)		
	Round 2	**Round 3**
Constant	-0.251	-0.671
	(0.813)	(0.832)
Group choice	0.656**	0.829***
	(0.197)	(0.199)

*** Sig. at 1%; ** Sig. at 5%; * Sig. at 10%

Table C.7. Explaining the Change in contributions in Rounds 4-6

Dependent Variable: Contribution change ($x_{i,t+1} - x_{i,t}$)				
	Round 4	**Round 5**	**Round 6**	**Pooled**[a]
α (constant)	0.178	0.808**	−0.102	0.198
	(0.400)	(0.381)	(0.250)	(0.221)
γ_1 (dist. from mode)	0.244*	0.474***	−0.200	0.259**
	(0.132)	(0.116)	(0.136)	(0.120)
γ_2 (dist. from mean)	0.236**	0.082	0.0659	0.176*
	(0.117)	(0.128)	(0.0858)	(0.096)
γ_3 (dist. from median)	−0.395**	−0.496***	0.187	−0.382***
	(0.165)	(01.68)	(0.188)	(0.149)
γ_4 (M-C score)	0.0435	−0.124	0.007	−0.0193
	(0.400)	(0.079)	(0.0539)	(0.0388)

*** Sig. at 1%; ** Sig. at 5%; * Sig. at 10%; Observations pooled for rounds 4 to 6.

Table C.8. Responding to friends and strangers

Dependent Variable: $(x_{i,t+1} - x_{i,t})$	**Round 5** ($t = 4$)	**Round 6** ($t = 5$)	**Round 5**	**Round 6**
	(1)	(2)	(3)	(4)
Constant (α)	0.0588	0.0746	0.1836	0.2261
	(0.236)	(0.163)	(0.2903)	(0.1975)
Distance from friends (η_F)	0.108*	0.0696	–	–
	(0.0541)	(0.0537)		
Distance from strangers (η_S)	−0.0258	0.0107	−0.0504	−0.0251
	(0.0707)	(0.0514)	(0.0789)	(0.0577)
η_{F1}	–	–	0.0287	−0.0526
			(0.128)	(0.1064)
η_{F2}	–	–	−0.1483*	−0.1216*
			(0.0882)	(0.0661)

*** Sig. at 1%; ** Sig. at 5%; * Sig. at 10%

APPENDIX D

Appendix for Chapter 4

D.1. Selective Comments

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

Females reported:

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

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

Males reported:

- To me this means the employee is not treated fairly or simply does not feel comfortable in the work environment.
- Discrimination in salary and at the workplace, acceptance at workplace, promotion, acceptance into societies and journals, respect.
- Harassment/mistreatment from coworkers and unfair compensation when compared to the opposite sex

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

D.2. Figures and Tables

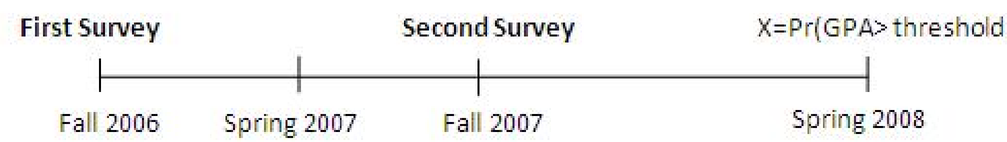

Figure D.1. Timeline of the surveys

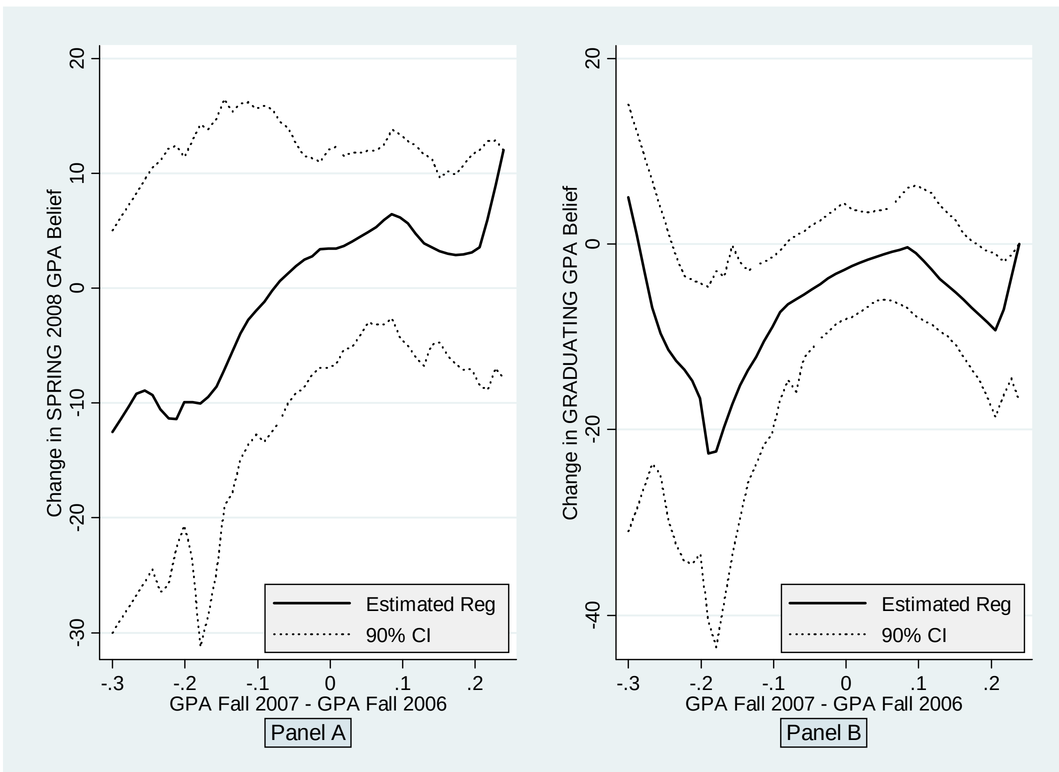

Figure D.2. Change in the GPA beliefs in response to change in GPA realized between the surveys.

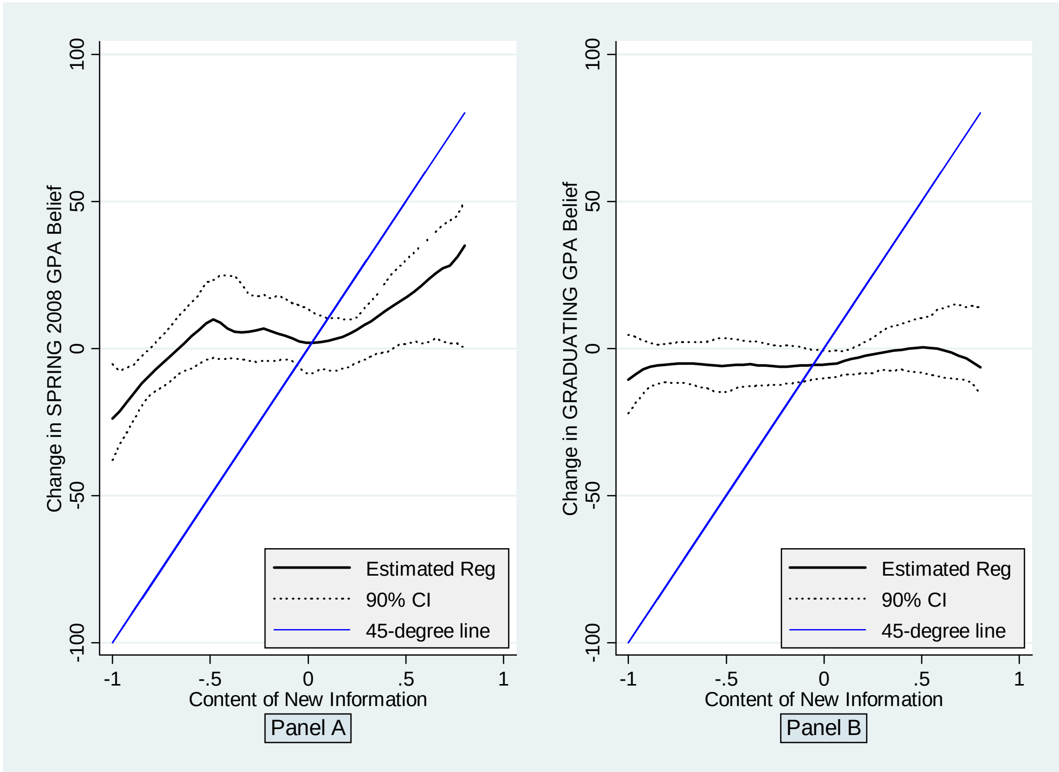

Figure D.3. Change in GPA beliefs in response to new information revealed between the two surveys.

Table D.1. Sample Characteristics

Characteristics	*Follow-up*[a] Freq.(Percent)	*Initial Survey*[b] Freq.(Percent)	*Population*[c] Freq.(Percent)
	(1)	(2)	(3)
Gender			
Male	51 (43.5)	69 (43)	465 (46)
Female	66 (56.5)	92 (57)	546 (54)
Total	117	161	1011
Ethnicity			
Caucasian	66 (56)	79 (49)	546 (54)
African American	10 (9)	11 (7)	71 (7)
Asian	35 (30)	56 (35)	232 (23)
Hispanic	1 (1)	5 (3)	61 (6)
Other	5 (4)	10 (6)	101 (10)
Declared Major?[d]			
Yes	61 (52)	90 (56)	182 (18)
No	56 (48)	71 (44)	829 (82)
Second Major?[e]			
Yes	55 (47)	78 (48.5)	–
No	62 (53)	83 (51.5)	–
Intl. Student?[f]			
Yes	5 (4)	8 (5)	40 (4)
No	112 (96)	153 (95)	971 (96)
Sec-Gen Immig.?[g]			
Yes	43 (37)	66 (41)	–
No	74 (63)	95 (59)	–
Average GPA[*]			
Male	3.51	3.48	3.26
Female	3.43	3.40	3.31

[a] Individuals who participated in the follow-up (second) survey
[b] Individuals who participated in the initial survey
[c] Population Statistics for sophomore class. (Source: Northwestern Office of Registrar)
[d] Whether the respondent has declared their major at the time of the INITIAL survey
[e] Whether the respondent was pursuing a second major in the INITIAL survey
[f] Whether the respondent is an international student
[g] Whether at least one of the respondent's parents is foreign-born, and the respondent was US-born
[*] Difference in GPAs within gender between the surveys is insignificant (2-tailed t-test)

Table D.2. Statistics on Average Annual Starting Salaries of Northwestern 2007 Graduates

Variable: Avg. Starting Salary of 2007 NU Graduates in each Category as reported by:	*Avg. Salary of Males*			*Avg. Salary of Females*			*% Diff in Salaries*[a]		
Category:	**Grad '07**[b]	**Males**[c]	**Females**[d]	**Grad '07**	**Males**	**Females**	**Grad '07**	**Males**	**Females**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Natural Sciences	36,286	42,690	50,340	33,200	38,347	45,422.5	8.50	10.17	9.38
	(11,898)	(17,028)	(17,614)	(12,785)	(15,930)	(16,093)		(8.14)	(8.93)
Math & Computer Sc.	68,000	48,929	51,667	54,000	45,857	49,667	20.59	5.73	5.02
	(12,550)	(11,958)	(49,667)	–	(10,523)	(18,016)		(7.34)	(7.76)
Social Sciences I	47,928	38,643	35,666	39,694	36,083	32,758	17.18	6.14	7.41
	(25,337)	(14,940)	(13,337)	(19,716)	(13,669)	(12,100)		(6.47)	(8.72)
Social Sciences II	57,877**	51,518	49,286	52,000	49,185	44,643	10.15	4.39	8.01
	(14,780)	(9,720)	(11,906)	(7,552)	(9,140)	(7,958)		(4.51)	(9.11)
Ethics and Values	31,667	32,900	35,231	25,000	32,500	32,461	2.11	1.45	7.31
	(2,886)	(6,118)	(12,001)	–	(6,587)	(10,316)		(7.44)	(9.04)
Area Studies	40,000	36,318	36,043	46,250	33,476	33,418	−15.62	7.21	6.79
	(12,062)	(11,838)	(12,062)	(15,654)	(11,199)	(11,052)		(8.74)	(7.68)
Literature & Fine Arts	30,142	31,601	32,345	36,800	30,601	29,814	−22.08	2.62	6.59
	(14,485)	(12,775)	(12,259)	(14,953)	(12,688)	(10,124)		(18.33)	(10.55)
Music Studies	30,000	–	30,000	30,400	–	30,000	−1.33	–	0
	(8,660)		–	(8,384.5)		–			
Educ & Social Policy	40,000	50,000	35,000	39,696	45,000	35,000	0.75	10.00	0
	(12,884)	–	–	(12,221)	–	–			
Communication Std	46,456.5	–	40,250	42,362	–	37,000	8.81	–	7.69
	(53,001)		(12,120)	(14,611)		(10,456)			(3.41)
Engineering	56,706*	50,053	54,732	52,671	49,103	51,429	7.11	2.73	5.85
	(10,211)	(15,448)	(20,962)	(12,309)	(21,314)	(20,913)		(18.87)	(11.42)
Journalism	47,875	30,000	30,000	35,694	35,000	30,000	2.54	−16.67	14.28
	(24,375)	–	–	(12,648)	–	–		–	–

Standard errors in parentheses; gender difference is significant at 10%; ** significant at 5%; *** significant at 1% (2-tailed T-test)

a The AVERAGE % difference between the female and male salary i.e. $\frac{\text{male salary - female salary}}{\text{male salary}} * 100$

b Average reported starting salary of a graduating member of the 2007 Class majoring in that category (Source: Northwestern Graduation Survey 2007)
c Response of Male survey respondents to the question: "What do you think was the average annual starting salary of Northwestern MALE graduates (of 2007) with Bachelor's Degrees in X.
d Response of Female survey respondents to the same question as in *c*

Table D.3. Explaining the errors in students' salary expectations

Dependent Variable: Log Absolute Error in Beliefs about:$^{\oplus}$	*Initial Survey*	*Starting Salaries for MALES*			*Starting Salaries for FEMALES*		
	All	**All**	**Overpred.**†	**Underpred.**‡	**All**	**Overpred.**	**Underpred.**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Major Declared[a]	−0.096	0.298	0.636***	0.117	0.130	0.363	0.122
	(0.182)	(0.187)	(0.281)	(0.273)	(0.168)	(0.286)	(0.213)
Cumulative GPA	0.189	−0.344*	−0.563*	−0.020	−0.235	−0.177	−0.197
	(0.235)	(0.206)	(0.309)	(0.283)	(0.233)	(0.326)	(0.332)
SAT Math	0.00033	0.0021**	0.0008	0.0042***	0.0034***	0.0030	0.0047***
	(0.0016)	(0.0011)	(0.0016)	(0.0015)	(0.0012)	(0.0017)	(0.0017)
SAT Verbal	−0.0013	0.00041	−0.018	0.00089	−0.0022*	−0.0059***	−0.00077
	(0.0014)	(0.0011)	(0.0016)	(0.0016)	(0.0013)	(0.0019)	(0.0015)
Female	0.212	0.365**	0.390*	0.259	0.404**	0.377*	0.271
	(0.190)	(0.158)	(0.211)	(0.219)	(0.161)	(0.221)	(0.213)
NU Credits	−0.0232	−0.010	0.048*	−0.058	0.014	0.063**	−0.0298
	(0.0292)	(0.0237)	(0.029)	(0.042)	(0.022)	(0.030)	(0.038)
Asian	−0.106	−0.115	−0.043	−0.408	−0.098	−0.0004	−0.305
	(0.280)	(0.248)	(0.360)	(0.322)	(0.301)	(0.409)	(0.375)
Foreign	−0.020	−0.123	−0.032	−0.110	−0.234	0.604	−0.380
	(0.569)	(0.314)	(0.819)	(0.304)	(0.366)	(0.545)	(0.399)
Sec-Gen Immigrant	0.361	0.163	−0.148	0.386	0.228	−0.047	0.359
	(0.254)	(0.193)	(0.287)	(0.260)	(0.258)	(0.350)	(0.316)
Studying Major[b]	−0.190	0.144	0.560*	−0.065	−0.105	−0.083	−0.060
	(0.161)	(0.148)	(0.294)	(0.187)	(0.186)	(0.392)	(0.198)
Studying Major × Decl.	−0.014	−0.288	−0.965**	0.343	−0.360	−0.047	−0.171
	(0.251)	(0.224)	(0.380)	(0.328)	(0.238)	(0.349)	(0.270)
Private High School	0.121	−0.093	−0.043	−0.136	−0.022	0.169	−0.173
	(0.156)	(0.164)	(0.252)	(0.204)	(0.161)	(0.240)	(0.192)
Low Parents' Income[c]	−0.153	0.131	0.475**	0.086	0.068	0.053	0.232
	(0.176)	(0.154)	(0.203)	(0.244)	(0.155)	(0.215)	(0.222)
Father went to College	−0.107	−0.616***	−0.776***	−0.387	−0.289	−1.150***	0.317
	(0.366)	(0.228)	(0.273)	(0.297)	(0.261)	(0.280)	(0.250)
Mother went to College	−0.638*	0.351*	0.177	0.174	0.277	−0.140	0.290
	(0.334)	(0.195)	(0.297)	(0.245)	(0.229)	(0.318)	(0.331)
Father studied major[d]	−0.205	0.044	0.397	−0.339	−0.332	−0.252	−0.376
	(0.241)	(0.207)	(0.279)	(0.347)	(0.227)	(0.406)	(0.259)
Mother studied major[e]	0.383**	−0.077	−0.029	−0.243	−0.194	−0.096	−0.281
	(0.192)	(0.167)	(0.239)	(0.274)	(0.205)	(0.321)	(0.285)
Resp. Random Eff.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	341	338	149	190	344	149	192
No. of Clusters	117	117	79	89	117	79	89

Notes for Table D.3

Estimates correspond to estimation of OLS model.
Cluster errors in parentheses; * sig at 10%; ** sig at 5%; *** sig at 1%

$\oplus$The dependent var, is the log of the absolute error: $\ln\left|\frac{\overline{s_{im}^{G}}-s_{obs\_m}^{G}}{s_{obs\_m}^{G}}\right|$ where $\overline{s_{im}^{G}}$ is the respondent's reported belief of the avg. salary of Northwestern 2007 grads of gender G in major m, and $s_{obs\_m}^{G}$ is the actual avg. salary earned by 2007 graduates in m.

† (‡) Sample restricted to observations where reported estimate is greater (less) than observed salary, i.e. $\overline{s_{im}^{G}} > s_{obs\_m}^{G}$ ($\overline{s_{im}^{G}} < s_{obs\_m}^{G}$)

a a dummy variable that equals one if the respondent had declared his/ her major at the time of the initial survey

b a dummy that equals one if the respondent's intended major category is same as category m in the salary question

c a dummy that equals one if parents' annual income is less that $150,000

d (e) a dummy that equals one if father's (mother's) field of study is the same as the salary question

Table D.4. Summary Stats about labor force participation, fertility beliefs and time use

Summary Statistics about labor force participation and fertility				
	At the age of 30		*At the age of 40*	
	Males	Females	Males	Females
Average beliefs of having:				
zero children	39.20	38.55	13.50	12.60
	(27.21)	(32.70)	(18.00)	(20.40)
one child	34.85	35.80	25.15	25.20
	(16.98)	(23.14)	(13.95)	(20.00)
two children	19.35	21.10	37.70	39.05
	(14.70)	(22.63)	(16.28)	(23.47)
three children	5.40	3.90	18.50	20.00
	(9.55)	(9.71)	(15.18)	(22.73)
four children	1.20	0.65	5.15	3.15
	(2.72)	(2.98)	(8.73)	(10.47)
Exp. number of Children	0.95	0.92	1.77	1.76
	(0.57)	(0.66)	(0.59)	(0.71)
Avg. beliefs of being:				
Full-time employed	91.75***	81.45	91.70***	81.01
	(8.80)	(19.10)	(8.80)	(21.40)
Part-time employed	6.03***	12.90	5.76***	12.67
	(7.51)	(13.71)	(6.99)	(13.86)
Not employed	2.20**	5.65	2.55**	6.31
	(3.45)	(9.80)	(3.76)	(11.36)
Full-time employed (Fall 2006)[a]	95.18	86.07	93.29	82.86
	(4.68)	(14.18)	((6.60)	(18.30)
Primary bread-earner[b]	76.40***	49.90	79.15***	47.15
	(15.90)	(19.35)	(15.55)	(18.90)
Fraction of time on house work[c]	0.24***	0.33	0.30***	0.39
	(0.11)	(0.13)	(0.12)	(0.15)

standard deviations in parentheses
*** (**) gender difference significant at 1% (5%) using the two-tailed t-test
[a] Avg. full-time labor force beliefs elicited in the initial (Fall 2006) survey
[b] Avg. belief (on a scale of 0-100) of being the primary bread-earner of the family
[c] Avg. time spent on house work (on a scale of 0-1)

Table D.5. Best Linear Predictor of Expectations of being active in the full-time labor force at 30

Dependent Variable: Belief of being active in the Full-Time Labor Force AT THE AGE OF 30						
	All	**Male**	**Female**	**All**	**Male**	**Female**
Cumulative GPA	−0.363	−0.021	7.78	−1.43	−0.93	6.53
	(3.51)	(3.96)	(6.05)	(3.45)	(4.11)	(6.09)
Foreign[a]	3.25	−6.14	53.44**	0.87	−7.59	45.98*
	(8.58)	(6.61)	(22.83)	(8.39)	(6.59)	(23.19)
Second-Generation Immigrant[b]	2.42	0.15	4.74	0.94	0.027	1.51
	(3.22)	(3.13)	(5.72)	(3.18)	(3.23)	(6.21)
SAT_Math[c]	−0.0313	−0.030	−0.060	−0.034	−0.034	−0.056
	(0.026)	(0.028)	(0.046)	(0.026)	(0.028)	(0.047)
SAT_Verbal	0.0036	0.010	0.012	−0.0043	0.0023	0.0015
	(0.026)	(0.029)	(0.039)	(0.026)	(0.030)	(0.041)
Female	−11.91***	omitted	omitted	−7.15*	omitted	omitted
	(3.19)			(3.97)		
Parents' Earnings:[d]						
$75,000-$150,000	−1.51	2.08	−4.86	−1.04	2.36	−5.24
	(4.12)	(3.97)	(6.97)	(4.09)	(3.96)	(7.08)
$150,000-$350,000	−4.05	−1.10***	−6.45	−3.98	−2.38	−5.45
	(4.22)	(4.18)	(6.92)	(4.14)	(4.17)	(7.10)
$350,000-$500,000	−20.23**	−6.40	−44.08***	−21.34***	−9.11	−42.85***
	(8.46)	(10.15)	(14.51)	(8.24)	(10.12)	14.52
> than $500,000	6.61	4.73	6.18	5.99	2.76	6.58
	(6.07)	(6.44)	(8.96)	(5.93)	(6.43)	(8.97)
Father went to College	6.59	−1.12	14.51	6.17	−0.38	14.07
	(6.58)	(6.17)	(11.66)	(6.43)	(6.15)	(11.65)
Mother went to College	−2.79	−8.57	0.637	−2.22	−7.51	−0.84
	(−5.32)	(6.83)	(8.42)	(5.29)	(6.83)	(8.81)
Stay-at-Home Mother	−6.24	−0.66	−5.96	−6.11	−2.77	−4.75
	(3.88)	(3.96)	(6.45)	(3.79)	(4.13)	(6.52)
Divorced	−3.63	−4.99	−3.84	−3.75	−5.07	−3.04
	(3.79)	(3.67)	(6.06)	(3.72)	(3.62)	(6.18)
Expected number of children[e]	–	–	–	−1.76	−0.20	−1.99
				(2.45)	(2.41)	(4.12)
Belief of primary bread-earner[f]	–	–	–	0.10	0.12	0.057
				(0.081)	(0.087)	(0.13)
Time spent on home production[g]	–	–	–	−0.25**	−0.21	−0.25
				(0.12)	(0.14)	(0.20)
R-Squared	0.2326	0.2019	0.2762	0.2949	0.2867	0.3200
No. of observations	117	51	66	117	51	66

Table D.6. Best Linear Predictor of Expectations of being active in the full-time labor force at 40

Dependent Variable: Belief of being active in the Full-Time Labor Force AT THE AGE OF 40						
	All	**Male**	**Female**	**All**	**Male**	**Female**
Cumulative GPA	1.34	0.12	8.61	0.11	−0.51	9.87
	(3.89)	(3.91)	(6.64)	(3.76)	(3.95)	(6.45)
Foreign[a]	7.72	−7.91	47.47*	7.82	−9.40	60.39**
	(9.50)	(6.53)	(25.07)	(9.14)	(6.29)	(25.19)
Second-Generation Immigrant[b]	4.61	1.02	4.88	2.78	−0.42	5.21
	(3.57)	(3.09)	(6.29)	(3.50)	(3.28)	(6.49)
SAT_Math[c]	−0.017	−0.030	−0.018	−0.016	−0.045	−0.014
	(0.029)	(0.027)	(0.051)	(0.028)	(0.027)	(0.049)
SAT_Verbal	0.0007	0.042	−0.004*	−0.011	0.029	−0.012
	(0.029)	(0.029)	(0.042)	(0.028)	(0.029)	(0.042)
Female	−11.94***	omitted	omitted	−1.46	omitted	omitted
	(3.54)			(4.61)		
Parents' Earnings:[d]						
$75,000-$150,000	−0.40	−0.039	−3.07	0.99	0.88	−2.59
	(4.56)	(3.92)	(7.65)	(4.39)	(3.80)	(7.36)
$150,000-$350,000	−5.07	−1.96	−8.80	−4.82	−3.82	−6.28
	(4.67)	(4.13)	(7.61)	(4.48)	(4.06)	(7.49)
$350,000-$500,000	−6.50	−14.79	−15.88	−8.91	−13.84	−26.46
	(9.37)	(10.03)	(15.94)	(9.02)	(9.79)	15.88
> than $500,000	9.92	3.12	11.93	9.73	1.79	11.29
	(6.72)	(6.37)	(9.83)	(6.50)	(6.24)	(9.52)
Father went to College	8.89	1.23	20.55	9.65	3.30	19.68
	(7.28)	(6.10)	(12.80)	(7.01)	(6.00)	(12.28)
Mother went to College	−5.90	−9.01	−6.55	−5.18	−8.37	−3.56
	(−5.89)	(6.75)	(9.24)	(5.68)	(6.48)	(9.13)
Stay-at-Home Mother	−9.91**	1.67	−17.04**	−9.86**	−0.41	−13.96**
	(4.30)	(3.92)	(7.08)	(4.13)	(3.94)	(6.89)
Divorced	2.94	−4.71	6.11	2.54	−5.13	−6.70
	(4.19)	(3.62)	(6.65)	(4.03)	(3.48)	(6.39)
Expected number of children[e]	–	–	–	−0.51	1.27	0.46
				(2.41)	(2.38)	(3.96)
Belief of primary bread-earner[f]	–	–	–	0.26***	0.16*	0.36**
				(0.09)	(0.093)	(0.14)
Time spent on home production[g]	–	–	–	−0.0021*	−0.002	−0.001
				(0.0013)	(0.001)	(0.002)
R-Squared	0.2198	0.2256	0.3057	0.3058	0.3483	0.3997
No. of observations	117	51	66	117	51	66

Table D.7. Perceptions of Monetary and Non-Monetary Discrimination

	Expected % Gap[a]				*% Females in class*[b]			*Males Poorly*[c]		*Fems Poorly*	
	At age of 30		At age of 40								
Category	Males[d]	Fems[e]	Males	Fems	Males	Fems	Recs[f]	Males	Fems	Males	Fems
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Natural Sciences	3.76	−1.97	7.86*	5.69	40.50	39.22	57.32	8.49	7.27	24.03	22.90
Math & Comp Sci	6.63	−3.94	5.95	−5.83	31.50**	25.22	34.12	7.17	6.53	23.71	29.19
Social Sciences I	1.28***	7.28	4.23	6.79	56.56*	60.30	61.72	8.55	11.56	12.71	14.75
Social Sciences II	6.10	−3.35	10.08	8.65	43.13	42.83	34.97	8.45	7.27	19.90	27.72
Ethics and Values	2.17	−5.21	0.95	5.16	55.39	55.98	39.18	9.56	11.01	12.07	15.71
Area Studies	−1.14**	6.50	−1.89**	6.15	59.84	58.15	77.27	10.52	9.87	11.07	13.19
Lit & Fine Arts	−0.55***	5.81	0.31*	5.75	64.82	66.16	73.11	11.19	11.22	8.94	13.15
Music Studies	−3.06***	6.17	–	–	59.25	57.22	50.97	13.05	10.63	9.25	13.16
Educ & Social Policy	3.27	1.96	–	–	66.21	68.86	76.24	13.47	16.57	9.82	13.18
Communication Std	3.14*	7.72	–	–	58.71	59.72	57.88	10.82	11.90	11.98	15.43
Engineering	7.21	6.53	2.31	−5.08	30.01	27.28	27.10	6.09	5.34	25.61	30.77
Journalism	−13.87*	9.62	–	–	58.90	60.22	71.42	10.47	11.69	12.03	16.30

*** gender difference significant (p-value < 0.01; two-tailed t-test); ** gender diff significant at the 5% level; * gender diff significant at the 1% level

[a] The AVERAGE % difference between the female and male salary i.e. $\frac{\text{male salary - female salary}}{\text{male salary}} * 100$

[b] Fraction of students in the major who are females (on a scale of 0-100)

[c] The average belief that males would be treated poorly in the jobs that would be available in each of the specified categories

[d(e)] Response of male (female) survey respondents to the relevant question

[f] The fraction of females amongst graduates with that major in 2005 and 2006 (source: Integrated Postsecondary Education Data System [IPEDS])

Table D.8. Reasons for gender differences in income within majors

Reasons for income differences between genders as reported by:	*Males*	*Females*
Characteristics and aptitudes actually differ between males and females	19.41**	11.36
	(24.30)	(14.87)
Different distribution of household duties	18.19	18.33
	(16.20)	(16.73)
Employers expect different characteristics between males and females	18.98**	27.77
	(18.35)	(21.79)
Employers' tastes given equal characteristics and household duties (i.e. discrimination on part of employers)	29.29	30.81
	(28.14)	(28.34)
Other Reasons	14.12	11.43
	(32.95)	(30.30)

Each cell is the AVERAGE contribution of the reason for the gender gap in salaries.
The exact question was: "How important do you think are the following explanations for the difference in salaries by gender? For this question you need to assign a number between 0 and 100 to each of the following explanations. Moreover, the responses SHOULD ALL SUM TO 100."
* gender difference is significant at 10%; ** significant at 5%; *** significant at 1% (2-tailed T-test)

Table D.9. Correlation patterns between various beliefs

	Females treated poorly	Males treated poorly	% Females in class	% Wage Gap at 30	% Wage Gap at 40	Enjoy Course-work
			Females			
Females treated poorly	1.00					
Males treated poorly	0.3997***	1.00				
% of Females in the class	−0.3523***	0.2191***	1.00			
% Wage Gap at the age of 30[a]	0.1114**	−0.0394	−0.0161	1.00		
% Wage Gap at the age of 40	0.1642***	−0.0191	−0.0450	0.6694***	1.00	
Enjoy coursework	−0.2616***	0.0926***	0.4654***	0.0880**	0.0307	1.00
Enjoy working at jobs	−0.1583***	−0.0205	0.2699***	0.1055**	0.0557	0.6196***
			Males			
Females treated poorly	1.00					
Males treated poorly	0.5575***	1.00				
% of Females in the class	−0.3584***	0.0692	1.00			
% Wage Gap at the age of 30	0.3418***	0.0767	−0.3108***	1.00		
% Wage Gap at the age of 40	0.3868***	0.1478**	−0.2279***	0.5647***	1.00	
Enjoy coursework	−0.1604***	−0.0483	0.0832*	−0.1704***	−0.2120***	1.00
Enjoy working at jobs	−0.0741	−0.0915*	0.0293	−0.1081**	−0.1038**	0.6704***

correlation significant using the Spearman's rank correlation coefficient: *** sig at 1%, ** sig. at 5%, * sig. at 1%

[a] The AVERAGE %difference between the female and male salary i.e. $\frac{\text{male salary - female salary}}{\text{male salary}} * 100$

Table D.10. Single Major Choice- Estimation of Preferences

	Original Model			*Extended Model*		
	All	Males	Females	All	Males	Females
Δu_1 for graduating within 4 years	0.451	0.334	0.236	0.440	0.424	0.315
	(0.635)	(0.756)	(0.939)	(0.637)	(0.829)	(0.952)
Δu_2 for graduating with a GPA of ≥ 3.5	0.781*	0.360	1.27*	0.784*	0.351	1.29*
	(0.434)	(0.481)	(0.679)	(0.435)	(0.451)	(0.662)
Δu_3 for enjoying the coursework	3.21***	2.67***	3.63***	3.23***	2.74***	3.60***
	(0.372)	(0.646)	(0.438)	(0.374)	(0.660)	(0.437)
γ_1 for hours/week spent on coursework^a	−0.001	0.0056	−0.0017	−0.001	0.010	−0.0032
	(0.011)	(0.017)	(0.014)	(0.011)	(0.016)	(0.014)
Δu_4 for approval of parents and family	1.90**	2.30***	1.49**	1.89***	2.36***	1.48**
	(0.421)	(0.576)	(0.647)	(0.420)	(0.571)	(0.623)
Δu_5 for finding a job upon graduation	−0.229	−0.479	0.023	−0.259	−0.642	−0.044
	(0.416)	(0.552)	(0.579)	(0.419)	(0.543)	(0.589)
Δu_6 for enjoying work at the available jobs	1.62***	0.550	2.47***	1.63***	0.517	2.46***
	(0.357)	(0.605)	(0.516)	(0.359)	(0.607)	(0.507)
Δu_7 for reconciling family and work at jobs	0.602	1.00	0.451	0.596	1.27*	0.510
	(0.401)	(0.701)	(0.522)	(0.399)	(0.728)	(0.542)
γ_2 for hours/week spent at work^b	0.0035	0.0022	0.0091	0.0035	0.0029	0.0082
	(0.0088)	(0.015)	(0.010)	(0.0089)	(0.015)	(0.0097)
γ_3 for the social status of the available jobs^c	0.897***	1.95***	0.151	0.880***	2.00***	0.187
	(0.298)	(0.489)	(0.380)	(0.299)	(0.459)	(0.381)
γ_4 for expected Income at the age of 30	−8.67e-08	3.33e-07	6.97e-07	−6.87e-08	7.56e-07	8.43e-07
	(1.07e-06)	(3.41e-07)	(1.00e-06)	(1.06e-06)	(3.36e-06)	(8.93e-07)
Δu_8 for females being treated poorly at jobs	–	–	–	0.245	0.850	−0.427
				(0.477)	(0.767)	(0.637)
Δu_9 for males being treated poorly at jobs	–	–	–	0.031	2.18*	−1.52
				(1.13)	(1.27)	(1.71)
Wald χ^2	243.85	130.99	196.86	245.67	144.90	197.19
Log-Likelihood	−989.27	−439.58	−523.19	−989.10	−436.44	−521.74
No. of Observations	117	51	66	117	51	66

† Estimates correspond to the estimation of a logit model on stated preference data
* significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses
a (b) - number of hours spent per week on coursework (job) varies between 0 and 100;
c - social status is on a scale of 1-8 (8 being the highest social status); normalized to be between 0.1-0.8
all other variables (except income) are probabilities between 0 and 1

Table D.11. Decomposition Analysis

	All	**Males**	**Fems**	**All**	**Males**	**Fems**
	(1)	(2)	(3)	(4)	(5)	(6)
Attributed to:						
Pecuniary Attributes[a]	17.55%	44.45%	8.90%	16.20%	41.40%	10.25%
Non-Pecuniary Attributes[b]	82.45%	55.55%	91.10%	80.60%	50.55%	87.35%
Beliefs about males and females treated poorly	–	–	–	3.20%	8.05%	2.40%
Attributed to:						
Parents' Approval + Enjoying Coursework	53.35%	45.55%	45.60%	52.40%	46.70%	45.15%
Coursework hrs/wk + GPA + Graduating in 4 yrs	13.35%	3.70%	17.30%	12.85%	22.70%	15.15%
Finding job + Job hrs/wk + Inc at 30 + Job Status	19.80%	41.75%	13.55%	18.95%	21.30%	13.40%
Reconcile work & family + Enjoying Work	13.50%	9.00%	23.55%	13.45%	5.35%	24.30%
Males and Females treated poorly at the jobs	–	–	–	2.30%	4.00%	2.00%

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 D.12. Stated reasons for choosing a major

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

* gender difference is significant at 10%; ** significant at 5%; *** significant at 1% (2-tailed T-test)
Each cell is the AVERAGE contribution of the reasons for the choice of majors
The exact question was: "In deciding your major, how important to you was each of the following reasons? For this question you need to assign an integer between 0 and 100 to each of the following reasons. Moreover, the responses SHOULD ALL SUM TO 100."

Table D.13. The change in beliefs over time

	Dropped Major[a]	**Least Pref. Major**	**Next Pref. Major**[b]	**Second Major**[c]	**Current Major**
Graduate in 4 years	-8.5	-0.30	-0.97	4.43	1.71
	(21.59)	(24.85)	(17.95)	(9.67)	(11.57)
	[71.43%]	[56.31%]	[72.41%]	[83.05%]	[90.27%]
GPA of ≥ 3.5	2.57	-6.66	-5.24	-1.60	-5.32
	(20.06)	(21.07)	(19.81)	(19.00)	(21.36)
	[57.14%]	[51.46%]	[44.83%]	[64.41%]	[51.33%]
Enjoy Coursework	-10.21	-9.15	-2.91	-0.26	-4.09
	(17.18)	(23.93)	(22.41)	(20.65)	(15.62)
	[50.00%]	[47.57%]	[51.72%]	[55.93%]	[56.64%]
Coursework hrs/week	-8.21	0.26	-3.02	-4.29	-5.49
	(13.95)	(13.47)	(12.12)	(16.15)	(13.28)
	[35.71%]	[53.40%]	[50.00%]	[47.46%]	[42.48%]
Approval of Parents	-2.64	-0.26	-4.48	-0.45	0.42
	(13.85)	(21.95)	(21.37)	(23.23)	(15.65)
	[71.43%]	[59.22%]	[56.90%]	[52.54%]	[66.37%]
Finding a job	2.29	-2.39	-2.31	-1.41	-1.19
	(27.87)	(24.77)	(24.13)	(20.66)	(23.52)
	[64.29%]	[51.46%]	[56.90%]	[59.32%]	[53.10%]
Enjoying work at jobs	-7.78	-12.66	-3.97	-5.98	-4.51
	(27.61)	(23.88)	(23.14)	(20.02)	(18.22)
	[35.71%]	[33.98%]	[48.28%]	[47.46%]	[57.52%]
Reconciling work & family	2.21	0.78	5.07	3.26	2.26
	(27.54)	(22.87)	(24.25)	(17.74)	(12.02)
	[35.71%]	[38.83%]	[55.17%]	[61.02%]	[52.21%]
Job hrs/week	0	2.69	1.72	1.84	2.26
	(11.94)	(11.29)	(19.12)	(12.75)	(12.02)
	[78.57%]	[57.28%]	[53.45%]	[57.63%]	[49.56%]
Salary at the age of 30	-4928.57	-4757.99	-100	22878.97	12856.38
	(98235.56)	(42834.78)	(33842.76)	(109003.7)	(55481.76)
	[7.14%]	[7.77%]	[20.69%]	[18.64%]	[7.96%]
Salary at the age of 40	27142.86	-9483.33	-1168.10	26309.57	-9711.28
	(81776.3)	(149714.9)	(80314.17)	(110969)	(161556.7)
	[0%]	[12.62%]	[15.52%]	[6.78%]	[6.19%]
No. of Observations	14	103	58	59	117

Standard errors in parentheses. * sig at 10%; ** sig at 5%; *** sig at 1%

[.] proportion of respondents for whom change in beliefs is $\leqslant$10 for binary outcomes; $\leqslant$ 5 for hrs/week; $\leqslant$1000 for income

a This is a major that the individual had once pursued;

b The second most preferred major for individuals without a second major;

c The individual's second major

Table D.14. The nature of evolution of beliefs over time

Dependent Variable: The posterior belief (i.e. belief in the follow-up survey)						
Depend. Var:		**Dropped Major**[a]	**Lest Pref. Mj**	**Next Pref. Mj**[b]	**Second Major**[c]	**Current Major**
Grad in 4 yrs	η	−120.49*(60.34)	48.48***(8.20)	24.82(15.05)	85.40***(7.75)	37.10***(10.24)
	γ	2.20***(0.64)	0.41***(0.096)	0.72***(0.16)	0.13(0.083)	0.63***(0.11)
	R	−0.55	1.46	0.39	6.97	0.60
GPA of more than 3.5	η	17.29(18.38)	0.46(4.81)	16.43**(7.70)	31.85***(10.66)	27.79***(6.71)
	γ	0.79***(0.24)	0.88***(0.073)	0.69***(0.10)	0.58***(0.13)	0.56***(0.085)
	R	0.25	0.14	0.44	0.72	0.77
Enjoying Coursework	η	−10.14(13.74)	11.13***(3.85)	38.61***(8.30)	49.92***(9.85)	49.96***(7.75)
	γ	0.99***(0.18)	0.53***(0.076)	0.43***(0.11)	0.37***(0.12)	0.37***(0.090)
	R	0.001	0.89	1.33	1.71	1.73
Coursework hrs/week	η	5.68(8.64)	6.82**(2.98)	10.63***(3.70)	0.97(4.45)	4.19(2.78)
	γ	0.45(0.31)	0.74***(0.11)	0.46***(0.14)	0.77***(0.17)	0.62***(0.10)
	R	1.22	0.36	1.19	0.29	0.61
Approval of Parents	η	−20.48(19.00)	10.95**(5.03)	16.56(12.42)	27.78***(10.24)	27.99***(8.68)
	γ	1.21***(0.22)	0.82***(0.073)	0.74***(0.15)	0.63***(0.13)	0.68***(0.099)
	R	−0.17	0.22	0.35	0.59	0.47
Finding a job	η	40.65*(21.08)	21.75***(6.36)	26.66***(8.53)	23.45***(8.41)	42.13***(7.35)
	γ	0.46(0.28)	0.62***(0.094)	0.58***(0.13)	0.63***(0.12)	0.41***(0.096)
	R	1.17	0.62	0.72	0.59	1.42
Enjoying work at jobs	η	25.77(19.90)	17.09***(5.30)	50.35***(11.42)	41.69***(9.56)	41.61***(7.74)
	γ	0.48(0.29)	0.41***(0.096)	0.26*(0.15)	0.35***(0.13)	0.43***(0.094)
	R	1.08	1.41	2.79	1.84	1.34
Reconciling work & family	η	31.47(43.81)	32.21***(7.22)	49.08***(9.56)	50.99***(6.91)	44.23***(5.88)
	γ	0.61(0.58)	0.52***(0.11)	0.36***(0.13)	0.34***(0.093)	0.39***(0.081)
	R	0.65	0.92	1.77	1.94	1.52
Job hrs/week	η	25.86**(11.07)	16.31***(5.01)	53.57***(8.20)	27.09***(7.05)	29.34***(3.76)
	γ	0.48**(0.22)	0.69***(0.11)	−0.11(0.17)	0.45***(0.15)	0.45***(0.075)
	R	1.10	0.45	−9.95	1.23	1.25
Salary at age 30	η	67.06**(22.30)	34.42***(4.83)	23.65**(9.57)	48.17*(28.34)	40.49***(9.88)
	γ	0.32**(0.15)	0.46***(0.052)	0.66***(0.12)	0.65*(0.34)	0.69***(0.093)
	R	2.13	1.19	0.51	0.54	0.44

Standard errors in parentheses. * sig at 10%; ** sig at 5%; *** sig at 1%

η is the constant; γ is coeff. on initial belief; R is importance of new information

a This is a major that the individual had once pursued;

b The second most preferred major for individuals without a second major;

c The individual's second major

Table D.15. Patterns in experimentation with majors

	All	Males	Females
Number of experimentations:			
Zero	45	22	23
One	44	19	25
Two or more	28	10	18
Total	117	51	66

Table D.16. Summary Statistics for experimentation with majors

Experimented with:	One Major		At least 2 Majors			
			First Major		*Second Major*	
Expected change in GPA[a]	0.17		−0.021		0.27	
	(0.92)		(0.97)		(1.03)	
Fraction who improve their GPA[b]	47.70%		35.71%		42.86%	
Number of courses in initial major[c]	3.02		2.44		1.81	
	(1.76)		(1.45)		(0.74)	
Reasons for dropping majors:[d]	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*
The initial major was too challenging	9.6	20.00	25.83	12.00	12.22	7.30
	(16.70)	(28.04)	(28.34)	(19.74)	(20.38)	(13.40)
The initial major was too easy	0.6	3.25	1.67	1.00	3.61	8.00
	(3)	(8.69)	(5.14)	(3.16)	(10.55)	(12.06)
I did not find the major interesting any more	29.8	29.74	20.5	16.00	31.61	23.20
	(33.11)	(23.30)	(19.43)	(20.79)	(24.00)	(26.61)
I got interested in something else	30.6	28.94	33.22	54.00	32.94	45.00
	(33.14)	(25.58)	(27.22)	(34.71)	(28.80)	(34.32)
My parents wanted me to change majors	0.8	0.79	2.67	1.00	2.39	1.00
	(2.77)	(2.51)	(7.99)	(3.16)	(7.99)	(3.16)
There was peer pressure to change majors	0.4	1.32	0.55	1.00	3.33	1.00
	(2)	(4.67)	(2.35)	(3.16)	(11.88)	(3.16)
Others	28.20	16.05	15.55	15.00	13.89	14.50
	(38.69)	(28.02)	(28.79)	(27.99)	(33.46)	(25.65)

[a] The mean change in GPA between the initial major and the next major pursued

[b] Fraction of individuals for whom their GPA in the new major increases (relative to previous major)

[c] Average number of courses the individual took in the initial major

[d] Each cell is the AVERAGE contribution of the reason for not continuing with the major. Individuals were asked to assign an integer between 0-100 to each of the reasons such that their responses all summed to a 100