

Using Subjective Expectations Data in Choice Models*

Basit Zafar[†]

Abstract

A pervasive concern with the use of subjective data in choice models is that they are biased and endogenous. This paper specifically addresses the questions of whether: (1) individuals report meaningful beliefs, and (2) do cognitive biases affect the way respondents update and report their beliefs. For this purpose, I collect a unique panel dataset of Northwestern University undergraduates which contains their subjective expectations about major-specific outcomes for their chosen major as well as for other alternatives in their choice set. Comparison of subjective beliefs with objective measures suggests that students respond "meaningfully". I do not find evidence of cognitive biases systematically affecting the reporting of beliefs. By analyzing patterns of belief-updating, I can rule out cognitive dissonance being of relevance in the current setting. Moreover, there doesn't seem to be any systematic (non-classical) measurement error in the reporting of beliefs: I don't find systematic patterns in mental recall of previous responses, or in the extent of rounding in the reported beliefs for the various majors. Overall, the results paint a very favorable picture for the use of subjective expectations data in choice models.

JEL Codes: D8, I2, J1, J7

Keywords: college majors; expectations; cognitive biases; endogeneity

*I am indebted to Charles Manski and Christopher Taber for extremely helpful discussions and comments. I also thank Meta Brown, Yonatan Eyal, Carlos Madeira, Wilbert van der Klauuw, Bob Willis, and participants at the 2009 CIPREE Workshop on Subjective Expectations in Econometric Models, the Econometric Society 2009 North American Summer Meeting, Northwestern's Labor Lunch, and the Federal Reserve Bank of New York Brown Bag Lunch for feedback and suggestions. Financial Support from a Northwestern University Graduate Research Grant is gratefully acknowledged. I thank all those involved in providing the datasets used in this paper. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole. All errors that remain are mine.

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

1 Introduction

Understanding any decision made under uncertainty requires one to study how expectations and preferences are used to make the choice. In the absence of data on expectations, existing empirical studies make non-verifiable assumptions on expectations, assume individuals are rational and form expectations in the same way, and use choice data to infer decision rules conditional on the maintained assumptions about expectations. This approach is problematic because 1) there is little reason to think that individuals form their expectations in the same way,¹ 2) observed choices may be consistent with several combinations of expectations and preferences (Manski, 1993), and 3) the information-processing rule has varied considerably among studies of schooling behavior, and it's not clear which is the correct one to use (given that individuals may use idiosyncratic rules to form their beliefs), and different rules yield vastly different predictions (Buchinsky and Leslie, 2009).

A solution to this identification problem is to directly elicit subjective beliefs and incorporate them into choice models. However, despite the fact that economists have increasingly been collecting and describing subjective data in the last decade or so (Manski, 2004), few studies incorporate subjective data into choice models (Lochner, 2007; Bellemare, Kroger, and van Soest, 2008; Delavande, 2008; Zafar, 2009). The most commonly suggested explanation for this is that subjective data are endogenous. In particular, when estimating choice models that incorporate subjective expectations, the researcher needs to elicit beliefs for outcomes associated with the choice that the individual has made as well as for outcomes associated with the other options in the individual's choice set. One concern is that if the respondent is asked about his beliefs, he is likely to exaggerate them in order to rationalize his choice (cognitive dissonance). Other cognitive biases such as respondents making little mental effort in answering questions, the lack of attitudes existing in a coherent form, and social desirability may also affect the way in which individuals report their beliefs (Bertrand and Mullainathan, 2001). Some of these concerns have been studied in cross-sectional analyses of subjective beliefs (Manski, 2004, and references therein). However, studying issues such as cognitive dissonance requires the researcher to have data on how beliefs evolve over time for outcomes associated with the choice that the individual made as well as for outcomes associated with choices that the individual

¹In fact, Arcidiacono, Aucejo, Fang, and Spenner (2009), Stinebrickner and Stinebrickner (2008), and Niederle and Vesterlund (2007) find that expectations about performance differ systematically by race, ability, and gender, respectively.

did not choose- such data usually don't exist.²

In order to address these questions, I designed and conducted two surveys that elicited subjective expectations from Northwestern University undergraduates regarding the choice of major. The first survey, administered to students in the early part of their sophomore year, collected details on respondents' demographics and data relevant for the estimation of the choice model; these data were used to estimate a choice model of college majors (Zafar, 2009). The second survey, conducted about a year after the first, collected data on how individuals revise their beliefs for major-specific outcomes. The major-specific outcomes for which beliefs were elicited include both outcomes that are realized in college and those that are realized in the workplace. Examples of the former include graduating in 4 years, enjoying the coursework, and parents approving of the choice, while examples of the latter include outcomes such as finding a job upon graduation, and being able to reconcile work and family at the jobs.

Subjective expectations data, like most data, have measurement error. However, any systematic rounding would be worrisome if such data were used directly to estimate choice models as in Delavande (2008) and Zafar (2009). The analysis shows that individuals adopt similar rounding practices when reporting their beliefs for outcomes associated with the various majors in their choice set, implying that such data can be used to obtain unbiased estimates in choice models. Most students successfully recall their previous responses. This, combined with the empirical evidence that individuals revise their beliefs in *sensible* ways, allows me to rule out the existence of non-attitudes (i.e., attitudes not existing in a coherent form, and individuals reporting something meaningless when asked a question for which they lack an attitude). Since it's not possible to directly observe individuals' thinking, I cannot refute the argument that individuals don't reveal their true beliefs.³ However, I present indirect evidence that individuals report their true beliefs by showing that individuals give internally consistent and sensible responses.

A major concern with subjective data is that individuals could report attitudes that are consistent with their behavior, i.e., cognitive dissonance may affect the subjective data. Here, it would imply that if an individual never pursued a major, they would tell themselves that

²There are, however, longitudinal studies that find that beliefs respond to changes in one's environment in meaningful ways. Examples include Bernheim (1988), Dominitz (1998), Dominitz and Manski (2005), Hurd and McGarry (2002) and Lochner (2007) who study revisions to expectations about social security benefits, income, mutual-fund investments, survival, and arrest, respectively.

³This is, however, a concern not specific to expectations data, and also applies to other survey research.

they never liked it anyway. Therefore, one would observe *unfavorable* changes in beliefs for outcomes in majors that an individual never pursued, and similarly large *favorable* changes in beliefs for outcomes for the major that the individual has stuck with. This systematic biasing of beliefs would be especially problematic if subjective data were used to estimate choice models. The panel on subjective beliefs allows me to check if cognitive dissonance affects the students' beliefs. Analysis of changes in beliefs suggests that this is not a concern: Average changes in beliefs of outcomes in a student's least preferred major and current major are not too different from those in other major categories.

Though the data in this study comes from a relatively homogenous sample, the raw data reveals substantial heterogeneity. This questions the assumptions that are imposed on expectations in the literature. To assess the "validity" of the subjective data, in instances where objective measures exist, I compare them to realizations of previous cohorts. For example, in the case of expected salary in the various majors, responses match up well with objective realities and students seem to be aware of income differences across majors. On the whole, the results in this paper bode well for the use of subjective expectations in choice models.

The paper is organized as follows: Section 2 outlines the choice model framework and formalizes the different biases. Section 3 describes the sources of data used in this study. Section 4 provides some description of the raw data, empirically investigates the extent to which various cognitive biases affect subjective data, and assesses the accuracy of the data. Finally, Section 5 concludes.

2 Theoretical Framework

The choice decision that I consider in this paper is that of college major. However, the general framework would apply to any choice under uncertainty. Individual i is confronted with the decision to choose a college major from his 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 $U_{ik}(\mathbf{b}, \mathbf{d})$ from choosing major k . Utility is a function of a vector of choice-specific discrete outcomes \mathbf{b} and a vector of continuous outcomes \mathbf{d} .⁴ Examples of outcomes in \mathbf{b} include graduating from college within four years, gaining approval of parents, and ability to reconcile family and work at the jobs. Examples

⁴One could allow the utility to be a function of individual characteristics as well.

of outcomes in \mathbf{d} include future income and number of hours spent on coursework. Both vectors, \mathbf{b} and \mathbf{d} , are uncertain at the time of the choice, and individual i possesses subjective beliefs $P_{ik}(\mathbf{b}, \mathbf{d})$ about the outcomes associated with choice of major k for all $k \in C_i$.⁵ If an individual chooses major m , then standard revealed preference argument (assuming that indifference between alternatives occurs with zero probability) implies that:

$$m \equiv \arg \max_{k \in C_i} \int U_{ik}(\mathbf{b}, \mathbf{d}) dP_{ik}(\mathbf{b}, \mathbf{d}). \quad (1)$$

The goal is to infer the preference parameters from observed choices. However, the expectations of the individual about the choice-specific outcomes are also unknown. The most one can do is infer the decision rule conditional on the assumptions imposed on expectations. This would not be an issue if there were reasons to think that prevailing expectations assumptions are correct. However, not only has the information-processing rule varied considerably among studies of schooling behavior, but most assume that individuals form their expectations in the same way. First, there is little reason to think that individuals form their expectations in the same way. Second, different combinations of preferences and expectations may lead to the same choice (Manski, 2002). A solution to this identification problem is to elicit subjective beliefs directly from individuals. Because economists have been skeptical of using subjective data, few studies have used this approach in the estimation of choice models (see, for example, Bellemare, Kroger, and van Soest, 2008; Delavande, 2008; Zafar, 2009). One reason for this skepticism is that cognitive biases may severely affect the reporting of beliefs (Bertrand and Mullainathan, 2001). Since identification of the preference parameters in equation (1) requires the elicitation of the respondent's expectations regarding the major that he has chosen as well as expectations regarding the *other* majors in his choice set that he could have chosen, concerns about the validity of subjective data are further exacerbated. Below I discuss some of the cognitive problems in more detail.

For simplicity, I assume that utility is linear and separable in outcomes. In that case, $U_i(\mathbf{b}, \mathbf{d}) = \sum_r u_r(b_r) + \sum_q \gamma_q d_q + \varepsilon_{ikt}$, where $u_r(b_r)$ is the utility associated with the binary outcome b_r , γ_q is a constant for the continuous outcome d_q , and ε_{ik} is a random term. Equation

⁵The vectors \mathbf{b} and \mathbf{d} are the set of outcomes common to all majors. It is the joint probability distribution of these outcomes $P_{ik}(\mathbf{b}, \mathbf{d})$ which is indexed by major k .

(1) can now be written as:

$$m \equiv \arg \max_{k \in C_i} \left(\sum_r \int u_r(b_r) dP_{ik}(b_r) + \sum_q \gamma_q \int d_q dP_{ik}(d_q) + \varepsilon_{ik} \right).$$

The additive separability of the utility function implies that only the marginal distribution of beliefs about the outcomes enter the expected utility. For the binary outcomes ($\{b_r\}_r$):

$$\begin{aligned} \int u_r(b_r) dP_{ik}(b_r) &= P_{ik}(b_r = 1)u_r(b_r = 1) + [1 - P_{ik}(b_r = 1)]u_r(b_r = 0) \\ &= P_{ik}(b_r = 1)\Delta u_r + u_r(b_r = 0), \end{aligned}$$

where $\Delta u_r \equiv u_r(b_r = 1) - u_r(b_r = 0)$, i.e., it is the difference in utility between outcome b_r happening and not happening. The linearity assumption of the utility function implies that only the expected value of the continuous outcomes matters since $\int U_i(\mathbf{b}, \mathbf{d}) dP_{ik}(\mathbf{b}, \mathbf{d}) = U_i(\int \mathbf{b}, \mathbf{d} dP_{ikt}(\mathbf{b}, \mathbf{d}))$. Thus, for the continuous outcomes ($\{d_q\}_q$), $\int d_q dP_{ik}(d_q)$ equals $E_{ik}(d_q)$, the expected value of the outcome. The expected utility that individual i derives from choosing major m is:

$$\begin{aligned} U_{im}(\mathbf{b}, \mathbf{d}, \{P_{im}(b_r = 1)\}_r, \{E_{im}(d_q)\}_q) = \\ \sum_r P_{im}(b_r = 1)\Delta u_r + \sum_r u_r(b_r = 0) + \sum_q \gamma_q E_{im}(d_q) + \varepsilon_{im}. \end{aligned} \quad (2)$$

An individual i with subjective beliefs $\{P_{ikt}(b_r), P_{ikt}(d_q)\}_{r,q} \forall k \in C_i$ chooses major m with probability:

$$\begin{aligned} \Pr(m | \{P_{ikt}(b_r), E_{ikt}(d_q)\}_{r,q}; k \in C_i) = \\ \Pr \left(\begin{array}{l} \sum_r P_{im}(b_r = 1)\Delta u_r + \sum_q \gamma_q E_{im}(d_q) + \varepsilon_{im} \\ > \sum_r P_{ik}(b_r = 1)\Delta u_r + \sum_q \gamma_q E_{ik}(d_q) + \varepsilon_{ik} \end{array} \right) \end{aligned} \quad (3)$$

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

In equation (3), $\{\Delta u_r\}_r$ and $\{\gamma_q\}_q$ are the parameters of the utility function that need to be estimated; Δu_r is the change in utility from the occurrence of outcome b_r , while γ_q is the parameter in the utility function for the continuous outcome d_q . $\{P_{ik}(b_r = 1)\}_r$ and $\{E_{ik}(d_q)\}_q$ are elicited directly from the respondent $\forall k \in C_i$.

The first concern with regards to using subjective data in choice models is **cognitive dissonance** (Festinger, 1957). Cognitive dissonance implies that individuals report attitudes that are consistent with their behavior. In the current setting, this bias would imply that if an individual has chosen a major m , he is likely to upgrade beliefs of outcomes associated with that

major and to devalue beliefs about outcomes associated with other majors in his choice set. Let $P_{ik}(b_r = 1)$ denote i 's true belief about the likelihood of outcome $b_r \forall k \in C_i$, and $P_{ik}^*(b_r = 1)$ the reported value. For desirable (undesirable) outcomes, cognitive dissonance would imply that $P_{im}^*(b_r = 1)$ is greater (less) than $P_{im}(b_r = 1)$, and $P_{ik}^*(b_r = 1)$ is less (greater) than $P_{ik}(b_r = 1) \forall k \in C_i$ and $m \neq k$. Therefore, a consequence of this bias would be that the estimated coefficients (in this case, Δu_r) would be upward biased.

A second potential problem is respondents making **insufficient mental effort** when reporting their beliefs, in particular for outcomes associated with majors they did not choose. Subjective expectations data, like most data, have measurement error. Studies using subjective expectations have documented the fact that respondents tend to round to the nearest 5 when reporting their beliefs on a scale of zero to 100. However, the concern is that the respondent's beliefs for outcomes associated with rejected majors in his choice set may exhibit greater noise. This *systematic rounding* (noise) would be worrisome if such data were used to estimate choice models since it would yield biased estimates. The bias arising from lack of mental effort would depend on the rounding practice of the respondent.⁶

Other concerns with using subjective data include social desirability, i.e., respondents giving the socially acceptable response in order to avoid looking bad in front of the interviewer, and non-meaningful responses either because respondents have incorrect beliefs, have no incentive to report their true belief, or because of the lack of beliefs existing in a coherent form. The rest of the paper empirically explores the extent to which these issues plague the data.

3 Data

The data used in this study come from two surveys that were administered to a sample of students in Northwestern University's undergraduate class of 2009. The first survey was administered to students in the early part of their sophomore year over the period from November 2006 to February 2007. I denote this as the *Fall 2006* or *initial* survey for the empirical analy-

⁶Here, I do not deal with the issue of non-systematic measurement error in the reporting of beliefs. Non-systematic measurement error would lead to bias towards zero in the case of classical measurement error, and no bias if the respondent reports his best estimate based only on the noisy measure of the mismeasured variable itself (see Hyslop and Imbens, 2001, for details).

There are two alternatives to dealing with (non-systematic) measurement error in subjective data: 1) infer the respondent's rounding practice, and interpret the numerical responses as intervals (Manski and Molinari, 2008), and then conduct the statistical inference by treating the subjective data as interval data (Manski and Tamer, 2002), or 2) modelling the preferences and beliefs jointly. This approach has been used by Lochner (2007) and Bellemare et al. (2008). Such an approach is, however, not feasible in models incorporating beliefs for several outcomes (as in Delavande, 2008, and Zafar, 2009).

sis. Since Northwestern University requires students to officially declare their majors by the beginning of their junior year, the timing of the initial survey corresponds to the period when students are actively thinking about which major to choose. The second survey was administered to a subset of the initial survey-takers at the beginning of their junior year, when students had presumably settled on their final majors.⁷ The survey spanned the period from November 2007 to February 2008. I denote it as the *Fall 2007* or *follow-up* survey.

Respondents for the initial survey were recruited by flyers posted around campus and by e-mailing a sample of eligible sophomores whose e-mail addresses were provided by the Northwestern Office of the Registrar. Prospective participants were told that the survey was about the choice of college majors and that they would receive \$10 for completing the 45-minute electronic survey. Respondents were required to come to the Kellogg Experimental Laboratory to take the electronic survey.

A total of 161 sophomores took the first survey, 92 of whom were females. The 45-minute survey consisted of two 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 (see Zafar, 2009). 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.

Of the 161 respondents who took the initial survey, 156 agreed to be contacted for the follow-up. About a year after the first survey, individuals who gave 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 administered 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 1 shows the characteristics of individuals who took the follow-up survey. For comparison, characteristics of the initial sample and the actual sophomore population are shown in columns (2) and (3), respectively. Respondents who took the follow-up survey seem 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

⁷Students can still change their major after their sophomore year, but they have to go through a formal process to do so.

significant. Table 3 shows that the distribution of majors in the Weinberg College of Arts and Sciences (WCAS) for the students taking the two surveys is similar, suggesting no differential attrition by field of study. As shown in Table 1, students of Asian ethnicity are overrepresented in the survey samples (both in the initial and follow-up survey) relative to their population proportion. Survey-takers, especially males, have higher average GPAs than their population counterparts. However, for the purposes of this study since I am primarily interested in analyzing how beliefs change over time, it's the selection into the follow-up survey that would be of concern. Based on observables, I don't find any selection in who decides to take the follow-up survey.

The follow-up survey primarily focused on how individuals revise their beliefs about major-specific outcomes. While the initial survey elicited beliefs about outcomes associated with all majors in the individual's choice set (which could be 8 or 9 majors),⁸ the follow-up survey elicited beliefs for major-specific outcomes only 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 pursuing at the time of the follow-up survey (one's most preferred major or current major), 2) the individual's second major (or the second most preferred major at the time of the follow-up survey 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 at the time of the follow-up survey).

The set of major-specific outcomes consists of binary outcomes, \mathbf{b} , and continuous outcomes, \mathbf{d} . The vector \mathbf{b} includes the outcomes:

b_1 successfully completing (graduating) a field of study in 4 years

b_2 graduating with a GPA of at least 3.5 in the field of study⁹

b_3 enjoying the coursework

b_4 parents approve of the major

b_5 obtain an acceptable job immediately upon graduation

b_6 enjoy working at the jobs available after graduation

b_7 are able to reconcile work and family while at the available jobs

⁸The College of Arts and Sciences at Northwestern University consists of 41 majors. Similar majors were pooled together. Table 2 shows the categorization of majors.

⁹This outcome is meant to capture the student's belief about academic ability in a major. The cutoff of 3.5 for graduating GPA was arbitrary.

while the vector \mathbf{d} consists of:

d_1 hours per week spent on the coursework

d_2 hours per week spent working at the available jobs

d_3 social status of the available jobs¹⁰

d_4 income at the available jobs

The survey elicited the probability of the occurrence of the binary outcomes, i.e., $\{P_{ikt}(b_r = 1)\}_{r=1}^7$ and the expected value for the continuous outcomes, i.e., $\{E_{ikt}(b_q)\}_{q=1}^4$. As mentioned above, the initial survey elicited these beliefs for *all* majors in the individual's choice set, while the follow-up survey elicited them for three different major categories in the individual's choice set.

Questions eliciting the subjective probabilities of major-specific outcomes were based on the use of percentages. An advantage of asking probabilistic questions relative to approaches that employ a Likert scale or a simple binary response (yes/no or true/false) is that responses are interpersonally comparable and allow the respondent to express uncertainty (see Manski, 2004, for an overview of the literature on subjective expectations). 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. The wording of the introduction was similar to that in Delavande (2008). An excerpt of the survey containing the introduction and list of questions dealing with the major-specific outcomes is presented in the Appendix. The full survey questionnaires are available on request from the author.

4 Empirical Analysis

This section explores the empirical nature of the various issues mentioned in section 2. I first describe the data in some detail. Table 4 shows the mean belief for each of the eleven outcomes for the eight main major categories as reported in the initial survey. This table shows substantial variation in mean belief for the same outcome across the various major categories, indicating that students do perceive differences in the occurrence for these outcomes across majors. For example, the mean belief of being able to graduate in 4 years varies from about 0.80 (on a 0-1 scale) for Engineering and Math & Computer Studies to 0.94 for Social Sciences I. Similarly,

¹⁰Social status of available jobs was elicited as an ordinal ranking. In hindsight, this question should have been asked in terms of the probabilistic chance of obtaining a high-status job, since the ordinal ranking does not reveal the respondent's uncertainty about the outcome.

the mean belief about gaining parents' approval varies from a low of 0.57 for Literature and Fine Arts to a high of 0.88 for Natural Sciences. The mean belief of finding an acceptable job after graduation varies from a low of 0.53 for Literature & Fine Arts to a high of 0.81 for Social Sciences II.

The mean beliefs reported in Table 4 mask the heterogeneity in responses across respondents for the *same* outcome. To highlight the heterogeneity in beliefs in my relatively homogenous sample, I present the subjective belief distributions reported in both surveys for graduating with a GPA of at least 3.5 in one's current major and one's least preferred major (at the time of the second survey) in Table 5. The table shows that respondents use the entire scale from zero to 100. Respondents tend to round off their responses to the nearest 5, especially for answers not at the extremes. There is a 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 (Bruine de Bruin et al., 2000).¹¹ However, the 50% response is not the most frequent one in the majority of the cases. Over time, it seems that individuals tend to revise downward their beliefs for graduating with a GPA of at least 3.5 for both their current major as well as their least preferred major. For example, in the initial survey, nearly half of the respondents believed there was a greater than 80% chance of graduating with a GPA of at least 3.5 in their current major. In the follow-up survey, the fraction of respondents who believed that to be the case had dropped to about 30%.

Figures 1 and 2 present the raw subjective data in a different way. The top panel of Figure 1 presents the histogram and cumulative belief distribution about enjoying coursework in one's current major, while the bottom panel presents the corresponding distribution in one's least preferred major. As in the case of beliefs about graduating with a GPA of at least 3.5, students' beliefs exhibit substantial heterogeneity and they use the entire scale from zero to 100. As one would expect, the belief distribution in the case of the least preferred major is skewed left relative to the distribution for one's preferred major. However, in both cases, the belief distribution in the initial survey first order stochastically dominates the belief distribution in the final survey. Over time, students revise their beliefs for enjoying coursework downward not only for the least preferred major but for their most preferred major also. Figure 2 presents the belief distribution about expected salary at age 30. In this case, the final survey belief distribution first order

¹¹This is what Bertrand and Mullainathan (2001) term as lack of attitude in a coherent form.

stochastically dominates the distribution from the initial survey for both major categories; students revise their beliefs upward about expected income for both their most preferred major as well as least preferred major.

I next empirically check for the extent of the various biases.

Social Desirability: Response distortion due to social desirability was mitigated by making the questionnaires anonymous, and by making respondents answer the survey online so that they don't have to answer directly to an interviewer. When taking the initial survey, students were not aware that there could be a potential follow-up survey, and their contact information was collected at the end of the survey only if they wanted to be contacted for a follow-up.

Accuracy of Beliefs: An important issue with regards to the use of subjective expectations data is their "accuracy", i.e., how they compare with objective realities. This is important for several reasons. One, to understand how well-informed individuals are when making choices and, second, to see how well assumptions like rational expectations explain real decision makers. At least two different ways have been used to assess the validity of subjective expectations: (1) comparing elicited expectations with future realizations,¹² and (2) comparing elicited expectations with historical realizations.¹³ This study uses the second approach since I don't observe realizations for most outcomes about which expectations are elicited. Moreover, it is not possible to assess the accuracy of non-pecuniary outcomes such as approval of parents or enjoying coursework since no objective measures exist for these outcomes.

When evaluating how the subjective data compare with the various objective measures that I use, it should be pointed out that there are at least four legitimate reasons why respondents' expectations may be different from them. First, Northwestern University undergraduates are a very specific demographic and the comparison groups that I've used might not be appropriate. Second, respondents might think that future distributions for the event of interest will differ from the current (or past) ones. Third, respondents may have private information about themselves which justifies them having different expectations. Fourth, the objective measures correspond to outcomes for students who choose to pursue that major. In this study, since beliefs are elicited from an individual about the occurrence of the various outcomes in his current

¹²This is the approach used by, for example, Dominitz (1998) and Hurd and McGarry (2002). The former study finds that income expectations predict actual income realizations, while the latter study finds that subjective survival probabilities predict actual survival.

¹³This approach has been used, for example, by Fischhoff et al. (2000) who study teen expectations for several significant life events and conclude that they are sensible.

major as well as for other majors in his choice set which he considered but did not choose, using data on realizations of students who choose that major may not be the *correct* objective measure. However, since that is the only data available, I use them for comparison purposes.

Table 6 compares the mean belief about graduating with a GPA of at least 3.5 and about expected income at the age of 30 in the various majors with realizations of bachelor graduates from institutions that are similar to Northwestern University. Column (1a) of Table 6 shows the mean GPA by major category of bachelor graduates (of selective Doctoral/Research universities) in the 2001 Baccalaureate & Beyond Longitudinal Study (B&B 2001), and column (1b) ranks the majors according to their GPA. Columns (2a) and (2b) provide the survey respondents' mean belief of being able to graduate with a GPA of at least 3.5 and the ranking of the majors in this dimension, respectively. The relative ranks of majors according to their GPA are similar in my sample and the B&B 2001, suggesting that students are aware of the relative difficulty of the various majors. Columns (3)-(4) of the table report the corresponding statistics for expected income at the age of 30. The objective measure in this case is the 2003 average annual salary of 1993 college graduates of selective colleges (Carnegie code 4) from the B&B 1993/2003 Study. The relative ranking of majors by income reported by respondents are similar to that computed using the B&B sample, indicating that students correctly perceive income differences across majors.

Students are expected have more favorable beliefs about positive outcomes for their more preferred majors. Therefore, analyzing the level of beliefs would not be useful in discerning the extent of the various biases. One way to assess the meaningfulness of individuals' responses is to analyze how beliefs for various outcomes change between the two surveys. Table 7 reports the mean change in the belief of a given outcome (disaggregated by how the individual ranks the major). The table also reports the mean absolute change of beliefs in parentheses and the fraction of responses which have remain *unchanged* since the initial survey in square brackets.¹⁴ The mean change in beliefs for almost all the binary outcomes is less than 10. Changes in beliefs for outcomes such as graduating in 4 years and parents' approval for a given major are smaller than those for other binary outcomes. This should not be surprising since one would expect individuals to have fairly precise beliefs about the occurrence of these outcomes at the

¹⁴I consider the belief of an outcome to have remain *unchanged* if: (1) the absolute change in beliefs is less than 5 points (on a scale of 0-100) for binary outcomes; (2) the absolute change in beliefs is less than 5 for hrs/week spent on coursework or job; (3) the absolute change in beliefs for salary is less than \$5000.

time of the initial survey and, therefore, to give similar responses in the follow-up. Similarly, as one would expect, mean changes in beliefs for outcomes associated with a dropped major are larger in magnitude than the corresponding changes in other major categories.

Table 8 presents the information shown in Table 7 in a slightly different way. The change in beliefs for each outcome is regressed onto dummies for the different major categories (second preferred major, second major, dropped major, least preferred major). The coefficients show the direction and magnitude of the mean change in beliefs about the various outcomes for each of the majors. Mean changes in the current major are indicated in the estimate of the constant. It is interesting to see that, on average, students revise their beliefs of graduating with a GPA of more than 3.5, enjoying coursework and enjoying work downward in their current major. The sign on most of the dummies indicate that individuals answer meaningfully and update their beliefs in meaningful ways. For example, for the dropped major, individuals revise their beliefs downward for all outcomes except graduating with a GPA of more than 3.5 and work flexibility. Even though most of these coefficients are not statistically different from zero (partially because of small subsamples), the fact that most of them have the *expected* sign is reassuring.

Cognitive Dissonance: Cognitive dissonance would imply that individuals revise their beliefs to preserve the positions that they are most committed to, and in a way that is consistent with their behavior (Festinger, 1957). For example, Mullainathan and Washington (2008) find evidence of cognitive dissonance in political attitudes; they find that opinion ratings of politicians reported by people eligible to vote exhibit greater polarization than those of comparable ineligible. In the current context, cognitive dissonance would imply that one would observe larger unfavorable changes in beliefs between the two surveys for outcomes in majors that an individual never pursued, and similarly larger favorable (or at least, less unfavorable) changes in beliefs for outcomes for the major that the individual has stuck with. As mentioned in section 2, distortion of responses due to cognitive dissonance would cause model estimates to be upward biased. However, Table 7 shows that 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. In particular, the direction in which beliefs are revised are consistent across the major categories. This is further confirmed in Table 8 which shows that mean changes in beliefs for almost all outcomes are similar (in sign and magnitude) across the various major categories; the estimated parameters are not statistically different from zero for all major dummies for most

outcomes (however, this could be because of small subsamples in some cases; for example, there are only 14 observations in the Dropped Major category). Moreover, as depicted in Table 7, beliefs about enjoying coursework and enjoying working at the jobs are revised downward in all major categories, and not only for the least preferred major. This is reassuring since cognitive dissonance would have implied favorable revisions for outcomes associated with one's current major.

It should be pointed out that if, in the initial survey, students reported beliefs that were already consistent with their choices, then revisions in beliefs for various outcomes across different majors would be similar and analyzing temporal patterns in changes in beliefs would not be meaningful. Though this form of endogeneity bias cannot be ruled out, it is highly implausible in this particular context. One, when students were interviewed for the initial survey at the beginning of their sophomore year, their choice of college major was reversible since they could easily switch majors until the end of the sophomore year. Second, as reported in Table 1, only half of the students had officially declared their majors at the time of the initial survey, suggesting that students were still unsure about their fields of study.¹⁵

Insufficient Mental Effort: As mentioned in section 2, differential mental effort when reporting beliefs for outcomes for different majors would lead to biased model estimates. The last row of each panel in Table 7 reports the proportion of responses which are not a multiple of 5 in curly brackets. There appears to be a fair amount of rounding to the nearest fifth. Responses seem to be rounded more for a given outcome as one moves from the right-most column to the left, i.e., more responses are rounded for outcomes associated with the least preferred major than for the second most preferred major, which in turn has more responses rounded than for outcomes associated with the current major. While most data have measurement error associated with them, the nature of rounding here would be especially worrisome if such data were used to estimate decision models. Moreover, this kind of rounding practice would be consistent with individuals make less mental effort when reporting their attitudes for outcomes associated with majors that they've not pursued. However, closer analysis of the data reveals that if one excludes responses at the extremes (i.e., 0, 1, 2 and 98, 99, 100), the extent of rounding is similar across the various major categories. Several studies have documented the reporting of responses at one-percent intervals at the extremes (Dominitz and Manski, 1997; Manski, 2004).

¹⁵Separate analysis of the students who had declared their major at the time of the initial survey and those who had not yields similar results. Results are not reported here because of small sample sizes.

One would expect to see more responses at the extremes if there's less uncertainty associated with that outcome; this is plausible in the case of outcomes associated with one's actual major. Given that there is no evidence of systematic rounding on part of the respondents, it is safe to conclude that such data can be directly used for choice analysis.

One final way in which I assess whether respondents exert sufficient mental effort when reporting their beliefs is by checking if they are aware of how their response in the follow-up survey compares to their response in the initial survey. The idea behind this exercise is that if respondents don't exert enough effort in reporting their beliefs, they would not be able to successfully report how their beliefs have changed over time. More specifically, individuals were asked about their *perceptions* of how their beliefs had changed since the initial survey. The wording of the questions was as follows:

"This question asks you to recall your beliefs i.e. responses to the questions you answered about Major X in the previous survey. Try to recall your beliefs from a year ago about the various outcomes in Major X, and then report whether YOU THINK your current beliefs are HIGHER, LOWER, or ABOUT THE SAME as the old (a year-ago) beliefs."

The elicited beliefs for the various outcomes in the two different surveys tells us how the beliefs *actually* changed between the two surveys. Table 9 presents a matrix of perceived changes in beliefs versus actual changes in beliefs.¹⁶ Each cell shows the number of responses (out of 117) that fall in that category. Perfect recall of earlier responses would imply that all off-diagonal cells would be zero- that is not the case. However, the number of occurrences of *absolute* error in recall (i.e., an individual perceiving their response to have increased when in fact it decreased, and vice versa) are only a small fraction of total responses. For example, in the case of beliefs about graduating in 4 years, the number of such occurrences is zero for the current major and second major, and 7 for the least preferred major. More importantly, there is no systematic pattern in which individuals make errors. One would be concerned if more errors in recall were made when reporting beliefs for the least preferred major or second major relative to those for the current major since this could imply that individuals make less mental effort when reporting their beliefs for outcomes not associated with their own major. That

¹⁶For this table, I have pooled responses for the least preferred major and dropped major into a single category, and similarly responses for the second major and second preferred major into a single category.

is, however, not the case. For example, in the case of beliefs of finding a job, the number of *absolute* errors is 3, 10, and 13 for the least preferred major, the second major, and the current major respectively. Table 9, however, shows that reported beliefs *actually* changed for a large fraction of respondents who perceived no change in their beliefs. One possible explanation for this is that respondents tend to round their beliefs.

5 Conclusion

This paper investigates a very specific question: Can subjective expectations data be used in choice models? This question is motivated by recent empirical work that underscores the importance of expectations in situations that involve uncertain outcomes, in particular schooling choices (Cunha, Heckman, and Navarro, 2004). Economic models of schooling choices usually make assumptions about how students form expectations. This is problematic because different information-processing rules yield significantly different predictions about individuals' schooling choices (Buchinsky and Leslie, 2009). A solution to this problem is to directly elicit subjective expectations data from the individuals. Though economists have increasingly undertaken the task of collecting subjective expectations data (Manski, 2004), concerns still remain about their use in decision models. This paper specifically addresses these concerns by analyzing whether respondents report *meaningful* beliefs and by investigating the extent to which cognitive biases such as social desirability, cognitive dissonance and insufficient mental effort plague subjective data. For this purpose, I collect a unique panel dataset of Northwestern University undergraduates which contains their subjective expectations about various major-specific outcomes.

The results in this paper bode well for the use of subjective expectations. Comparison of elicited beliefs with objective measures (in this case, realizations of previous cohorts) reveals that students are aware of income differences across college majors as well as of differences in how academically challenging the various majors are, suggesting that students respond in meaningful ways. Analysis of how beliefs evolve over time reveals that biases like cognitive dissonance, systematic rounding, or non-attitudes do not confound the data.

This paper adds to the literature on the validity and extent of bias of self-reported survey data (Bound, Brown, and Mathiowetz, 2001; Buchinsky and Leslie, 2009). To date, there is little agreement in the literature as to the validity and unbiasedness of such data. The results in the paper show that such data is indeed meaningful and can be used successfully to understand

how individuals make choices. However, this paper is clearly not the last word on this subject and more studies of this nature need to be conducted, especially because the data used in this study come from a very stylized setting. Moreover, the particular setting is one where, at the time of the first survey, the student's decision of what major to choose was reversible at a low cost. One would expect the impact of biases on beliefs to be stronger in settings where the decision is irreversible (or more costly to reverse).

References

Arcidiacono, Peter, Esteban Aucejo, Hanming Fang, and Kenneth Spenner. 2009. "Does Affirmative Action Lead to Mismatch? A New Test and Evidence." Duke University Working Paper.

Bellemare, Charles, Sabine Kroger, and Arthur van Soest. 2008. "Measuring Inequality Aversion in Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica*, 76(4): 815-839.

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

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

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." Chapter in *Handbook of Econometrics*, V. 5, eds. Edward Leamer and James Heckman, 3705-3843. Elsevier Science.

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

Buchinsky, Moshe, and Phillip Leslie. 2009. "Educational Attainment and the Changing U.S. Wage Structure: Dynamic Implications on Young Individuals' Choices." Working Paper.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2004. "Separating Heterogeneity from Uncertainty in Life Cycle Earnings." *Oxford Economic Papers*, 57(2): 191-261.

- Delavande, Adeline. 2008. "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice." *International Economic Review*, 49(3): 999-1042.
- Dominitz, Jeff and Charles Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31: 1-26
- Dominitz Jeff. 1998. "Earning Expectations, Revisions, and Realizations." *Review of Economics and Statistics*, 80: 374-388
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Hurd, Michael, and Kathleen McGarry. 2002. "The Predictive Validity of Subjective Probabilities of Survival." *Economic Journal*, 112(482): 966-985.
- Hyslop, Dean, and Guido Imbens. 2001. "Bias From Classical and Other Forms of Measurement Error." *Journal of Business and Economics Statistics*, 19(4): 475-481.
- Lochner, Lance. 2007. "Individual Perceptions of the Criminal Justice System." *American Economic Review*, 97(1): 444-460.
- Manski, Charles. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" In *Studies of Supply and Demand in Higher Education*, ed. Charles Clotfelter and Michael Rothschild, 43-57. Chicago: University of Chicago Press.
- Manski, Charles. 2002. "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response." *European Economic Review*, 46(4-5): 880-891.
- Manski, Charles. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329-1376.
- Manski, Charles, and Francesca Molinari. 2008. "Rounding Probabilistic Expectations in Surveys." *Journal of Business and Economic Statistics*, Forthcoming.
- Manski, Charles, and Elie Tamer. 2002. "Inference on Regressions with Interval Data on a Regressor or Outcome." *Econometrica*, 70(2): 519-546.
- Mullainathan, Sendhil and Ebonya Washington. 2008. "Sticking with your Vote: Cognitive Dissonance and Political Attitudes." *American Economic Journal: Applied Economics*, Forthcoming.

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

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

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

6 Appendix

6.1 Survey Excerpt

The following 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.

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 MALE graduates (of 2007) with Bachelor's Degrees in Natural Sciences?

Q10 What do you think was the average annual starting salary of Northwestern FEMALE graduates (of 2007) with Bachelor's Degrees in Natural Sciences?

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

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

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

Q10 What kind of jobs are you thinking of?

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

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

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

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

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

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

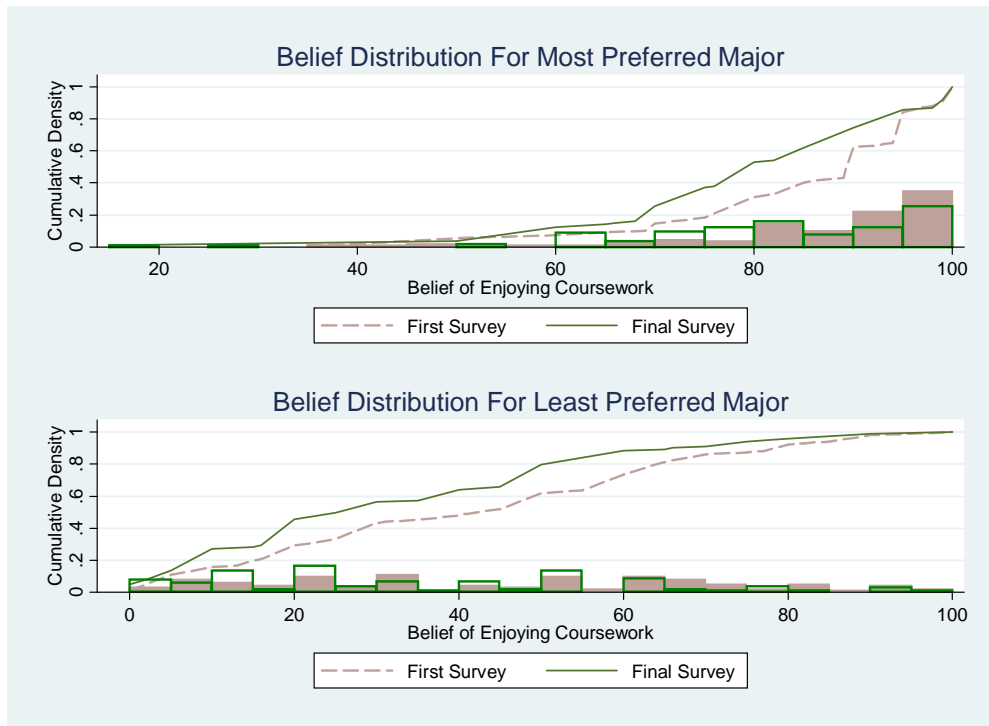


Figure 1: Belief Distribution of Enjoying Coursework.

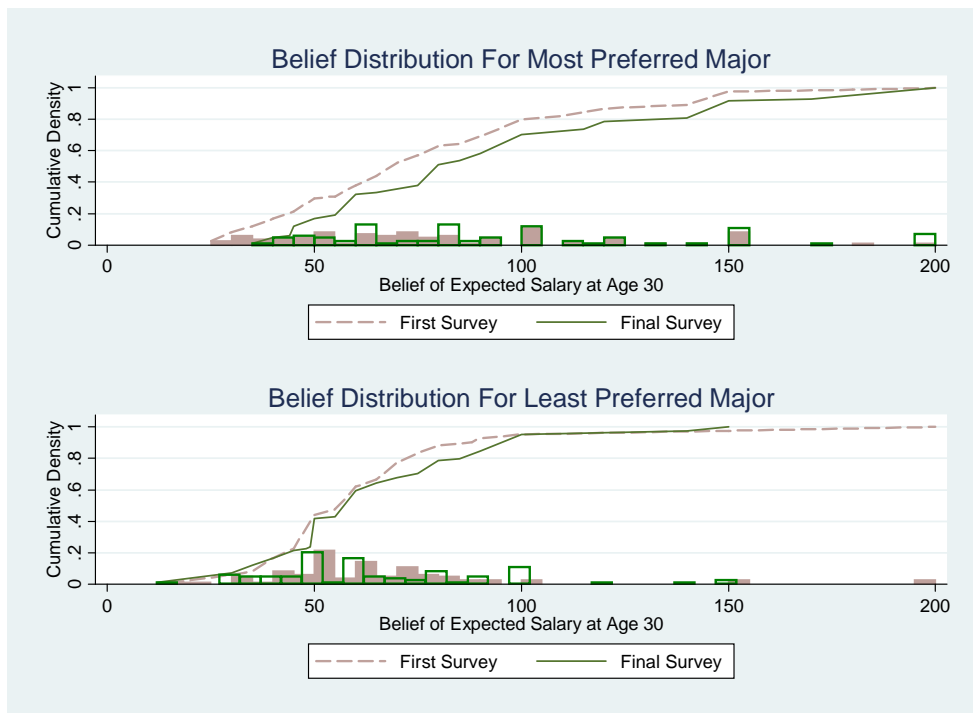


Figure 2: Belief Distribution of Expected Income at Age 30.

Table 1: Sample Characteristics

Characteristics	Follow-up Survey ^a	Initial Survey ^b	Population ^c
	Freq.(Percent)	Freq.(Percent)	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)	477 ^h (47)
No	56 (48)	71 (44)	534 (53)
Second Major?^e			
Yes	55 (47)	78 (48.5)	—
No	62 (53)	83 (51.5)	—
International Student?^f			
Yes	5 (4)	8 (5)	40 (4)
No	112 (96)	153 (95)	971 (96)
Second-Gen Immigrant?^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 the sophomore class. (Source: Northwestern Office of the Registrar)

^d Whether the respondent has declared a major at the time of the INITIAL survey

^e Whether the respondent was pursuing a second major at the time of 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 born in the U.S.

^h Statistic obtained from Registrar's Office at the end of the Fall 2006 quarter (during/middle of first survey)

* Difference in GPAs within gender between the two surveys is insignificant (2-tailed t-test)

Table 2: Categorization of Majors

Category Name	Majors in Category
WCAS Majors ⁺	
<i>a</i> Natural Sciences	Biological Sciences; Chemistry; Environmental Sciences; Geography*; Geological Sciences; Integrated Science; Materials Science; Physics
<i>b</i> Math & Computer Sciences	Cognitive Science; Computing and Information Systems; Mathematics; Statistics
<i>c</i> Social Sciences I	Anthropology; Gender Studies*; History; Linguistics; Political Science; Psychology; Sociology
<i>d</i> Social Sciences II	Economics; Mathematical Methods in the Social Sciences
<i>e</i> Ethics and Values	Legal Studies*; Philosophy; Religion; Science in Human Culture*
<i>f</i> Area Studies	African American Studies; American Studies; Asian and Middle East Languages & Civilization; European Studies; International Studies*; Slavic Languages and Literatures
<i>g</i> Literature & Fine Arts	Art History; Art Theory and Practice; Classics; Comparative Literary Studies; Drama; English; French; German; Italian; Spanish
Non-WCAS Majors	
<i>h</i> 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>i</i> Education and Social Policy ²	Human Development and Psychological Services; Learning and Organizational Change; Secondary Teaching; Social Policy
<i>j</i> Communication Studies ³	Communication Studies; Dance; Human Communication Science; Interdepartmental Studies; Performance Studies; Radio/Television/Film; Theater
<i>k</i> Engineering ⁴	Applied Mathematics; Biomedical Engineering; Chemical Engineering; Civil Engineering; Computer Engineering; Computer Science; Electrical Engineering; Environmental Engineering; Industrial Engineering; Manufacturing & Design Engineering; Materials Science & Engineering; Mechanical Engineering
<i>l</i> Journalism ⁵	Journalism
*Adjunct majors (these do not stand alone)	
+ Majors in the Weinberg College of Arts & Sciences (WCAS)	
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 3: Distribution of WCAS Majors in the Two Surveys

WCAS Majors ^a	Follow-up Survey ^b		Full Sample ^c		Initial Survey Sub-Sample ^d		Total Count ^e
	Freq	(%)	Freq	(%)	Freq	(%)	Freq
Natural Sciences	22	(19)	31	(19)	25	(21)	25
Math & Computer Sci.	2	(1.5)	4	(2.5)	4	(3.5)	6
Social Sciences I	33	(28)	41	(25.5)	33	(28)	44
Social Sciences II	35	(30)	48	(30)	32	(27.5)	42
Ethics and Values	1	(1)	4	(2.5)	-		2
Area Studies	8	(7)	13	(8)	8	(7)	18
Literature & Fine Arts	16	(13.5)	20	(12.5)	15	(13)	20
Total	117	(100)	161	(100)	117	(100)	157

^a Majors that appear in each category are listed in Table 2.

^b Major distribution of students in the Follow-up Survey. In cases where the survey respondent has more than one major in WCAS, only the first one is included.

^c Distribution of majors (first major only) at the time of initial survey of all 161 students.

^d Distribution of majors (first major only) in the initial survey of the 117 students who took the follow-up.

^e Distribution of majors (all WCAS majors of the student) in the initial survey of the 117 students who took the follow-up.

Table 4: Summary Statistics: Mean and Standard Deviation of Elicited Beliefs

	Graduate in 4 years	GPA ≥ 3.5	Enjoy Courses	Courses hrs/wk	Parents’ Approval	Find Job	Enjoy Work	Reconcile Family	Job hrs/wk	Status of Jobs	Income At 30
Natural Sciences	0.85 (0.20)	0.56 (0.28)	0.58 (0.26)	28.12 (12.10)	0.88 (0.15)	0.74 (0.20)	0.66 (0.24)	0.60 (0.22)	50.76 (12.44)	0.77 (0.17)	95.46 (82.89)
Math & Computer Sci	0.83 (0.21)	0.58 (0.27)	0.50 (0.26)	26.76 (12.33)	0.76 (0.22)	0.73 (0.19)	0.54 (0.23)	0.68 (0.20)	44.05 (10.04)	0.59 (0.17)	73.60 (51.48)
Social Sciences I	0.94 (0.08)	0.80 (0.16)	0.58 (0.26)	22.01 (10.29)	0.75 (0.23)	0.63 (0.20)	0.74 (0.19)	0.74 (0.16)	44.22 (9.65)	0.56 (0.16)	69.22 (38.27)
Social Sciences II	0.88 (0.13)	0.64 (0.25)	0.64 (0.23)	26.79 (12.43)	0.83 (0.16)	0.81 (0.14)	0.61 (0.23)	0.62 (0.20)	51.12 (11.75)	0.64 (0.18)	107.16 (82.63)
Ethics and Values	0.90 (0.12)	0.77 (0.17)	0.70 (0.18)	23.17 (11.85)	0.64 (0.26)	0.55 (0.21)	0.63 (0.19)	0.65 (0.20)	44.82 (11.26)	0.42 (0.19)	69.71 (42.25)
Area Studies	0.91 (0.11)	0.79 (0.18)	0.72 (0.20)	22.53 (10.43)	0.62 (0.26)	0.58 (0.20)	0.64 (0.19)	0.70 (0.16)	43.17 (9.38)	0.36 (0.17)	59.25 (26.32)
Literature & Fine Arts	0.91 (0.12)	0.78 (0.16)	0.70 (0.24)	23.98 (11.54)	0.57 (0.29)	0.53 (0.23)	0.64 (0.22)	0.76 (0.16)	41.09 (9.66)	0.35 (0.17)	54.62 (26.39)
Engineering	0.82 (0.21)	0.48 (0.27)	0.42 (0.27)	27.36 (11.65)	0.85 (0.17)	0.83 (0.15)	0.57 (0.22)	0.64 (0.18)	48.35 (10.32)	0.72 (0.16)	93.90 (64.30)

Binary outcomes (all outcomes except coursework hrs/wk, job hrs/wk, social status of jobs; income at 30) are on a 0-100 scale (and then divided by 100);

Coursework hrs/wk and job hrs/wk elicited on a scale of 0-70; Social status elicited on a 0-0.9 scale; Income at age 30 is expressed in 1000s.

There are 117 observations for each cell (except for outcomes related to Engineering for which there are 110 observations)

Table 5: Beliefs of Graduating with a GPA of at Least 3.5

Percent chance of graduating with a GPA ≥ 3.5 in:								
Reported in:	Current Major				Least Preferred Major			
	<u>Follow-up Survey</u>		<u>Initial Survey</u>		<u>Follow-up Survey</u>		<u>Initial Survey</u>	
Subj. Belief:	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
0	1	0.9	1	0.9	6	5.8	-	0
1	-	0.9	-	0.9	-	5.8	3	2.9
2	-	0.9	-	0.9	2	7.8	-	2.9
3	-	0.9	-	0.9	-	7.8	1	3.9
5	4	4.4	-	0.9	1	8.7	2	5.9
10	1	5.3	1	1.8	5	13.6	1	6.9
12	-	5.3	-	1.8	-	13.6	1	7.8
15	-	5.3	-	1.8	1	14.6	2	9.8
20	-	5.3	2	3.7	12	26.2	4	13.7
21	-	5.3	-	3.7	1	27.2	-	13.7
25	2	7.1	1	4.6	5	32.0	1	14.7
30	-	7.1	1	5.5	4	35.9	5	19.6
33	-	7.1	-	5.5	-	35.9	1	20.6
35	-	7.1	-	5.5	1	36.9	3	23.5
40	1	8.0	2	7.3	5	41.8	6	29.4
45	2	9.7	1	8.3	2	43.7	3	32.4
50	16	23.9	4	11.9	7	50.5	10	42.2
55	1	24.8	1	12.8	1	51.5	1	43.1
60	7	31.0	9	21.1	7	58.2	8	51.0
65	4	34.5	3	23.9	2	60.2	3	53.9
68	-	34.5	-	23.9	1	61.2	-	53.9
70	10	43.4	8	31.3	6	67.0	10	63.7
73	-	43.4	1	32.1	-	67.0	-	63.7
75	15	56.6	7	38.5	3	69.9	3	66.7
76	-	56.6	1	39.5	-	69.9	-	66.7
79	-	56.6	1	40.4	-	69.9	-	66.7
80	13	68.1	13	52.3	7	76.7	5	71.6
82	1	69.0	2	54.1	-	76.7	1	72.6
85	7	75.2	9	62.4	-	76.7	5	77.5
87	-	75.2	1	63.3	-	76.7	-	77.5
88	-	75.2	-	63.3	-	76.7	1	78.4
89	-	75.2	2	65.1	-	76.7	-	78.4
90	11	85.0	10	74.3	9	85.4	9	87.3
91	1	85.8	1	75.2	-	85.4	2	89.2
92	-	85.8	2	77.1	-	85.4	-	89.2
95	3	88.5	10	86.2	7	92.2	2	91.2
96	-	88.5	1	87.2	-	92.2	-	91.2
98	1	89.4	4	90.8	-	92.2	3	94.1
99	2	91.2	2	92.7	2	94.2	2	96.1
100	10	100	8	100	6	100	4	100

Table 6: Comparing GPA and Income Beliefs with Objective Measures

	Average GPA ^a		Belief GPA $\geq 3.5^b$		Average Salary ^c		Belief of Income at 30 ^d	
	Mean	Rank ^e	Mean	Rank	Mean	Rank	Mean	Rank
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Natural Sciences	3.22	5	0.56 (0.28)	7	75.86	3	95.46 (82.89)	2
Math & Computer Sci	3.21	6	0.58 (0.27)	6	73.32	4	73.60 (51.48)	4
Social Sciences I	3.29	1	0.80 (0.16)	1	72.73	5	69.22 (38.27)	6
Social Sciences II	3.09	8	0.64 (0.25)	5	78.10	2	107.16 (82.63)	1
Ethics and Values	3.29	1	0.77 (0.17)	4	68.23	6	69.71 (42.25)	5
Area Studies	3.29	1	0.79 (0.18)	2	68.23	6	59.25 (26.32)	7
Literature & Fine Arts	3.29	1	0.78 (0.16)	3	62.87	8	54.62 (26.39)	8
Engineering	3.11	7	0.48 (0.27)	8	89.26	1	93.90 (64.30)	3

^aMean GPA of bachelor graduates of Doctoral/Research Universities in 2001 (Source: 2001 Baccalaureate and Beyond Longitudinal Study).

^bBelief of survey respondents about graduating with a GPA ≥ 3.5 (on a scale of 0-100) and divided by 100. (N=117 except for Engineering which has 110 observations).

^cAverage salary in 1000s (in 2007 dollars) in 2003 of college graduates of 1993. Restricted to selective Doctoral/Research Universities with Carnegie Code 4 (Source: 1993/03 Baccalaureate and Beyond Longitudinal Study).

^dExpected salary at the age of 30 (in 1000s) elicited from survey respondents.

^eMajors are ranked from highest GPA or expected salary (rank 1) to lowest (rank 8).

Table 7: Summary Statistics about Changes in Beliefs

	Dropped Major^a	Least Pref. Major^b	Next Pref. Major^c	Sec Major^d	Current Major
Graduate in 4 years	-8.5 (12.5) [7.69%] {1.7%}	-0.30 (15.85) [37.61%] {7.7%}	-0.97 (9.65) [28.20%] {7.7%}	4.43 (6.47) [32.48%] {6.0%}	1.71 (4.75) [75.21%] {17.1%}
GPA of ≥ 3.5	2.57 (13.71) [5.98%] {0.9%}	-6.66 (16.09) [29.91%] {5.1%}	-5.24 (16.37) [11.96%] {0.9%}	-1.60 (12.26) [23.08%] {4.3%}	-5.32 (14.99) [33.33%] {4.3%}
Enjoy Coursework	-10.21 (14.35) [3.41%] {0%}	-9.15 (18.97) [23.93%] {5.1%}	-2.91 (16.43) [15.38%] {1.7%}	-0.26 (14.74) [18.80%] {2.6%}	-4.09 (11.89) [35.04%] {9.4%}
Coursework hrs/week	-8.21 (12.07) [4.27%] {2.6%}	0.26 (9.32) [47.00%] {12.8%}	-3.02 (8.63) [24.79%] {6.0%}	-4.29 (10.50) [23.93%] {16.2%}	-5.49 (9.95) [41.02%] {22.2%}
Approval of Parents	-2.64 (8.21) [6.84%] {1.7%}	-0.26 (14.37) [35.04%] {1.7%}	-4.48 (15.28) [16.24%] {2.6%}	-0.45 (15.79) [18.80%] {5.1%}	0.42 (10.13) [48.71%] {8.5%}
Finding a job	2.29 (18.00) [5.13%] {0.9%}	-2.39 (18.33) [29.91%] {3.4%}	-2.31 (16.17) [21.37%] {0.9%}	-1.41 (15.00) [17.95%] {3.4%}	-1.19 (16.16) [31.62%] {5.1%}
Enjoying work at jobs	-7.78 (23.35) [2.56%] {0.9%}	-12.66 (21.38) [17.95%] {2.6%}	-3.97 (17.59) [16.24%] {0.9%}	-5.98 (16.22) [12.82%] {1.7%}	-4.51 (13.19) [39.31%] {3.4%}
Reconcile work & family	2.21 (20.07) [2.56%] {1.7%}	0.78 (18.33) [17.94%] {0.9%}	5.07 (16.38) [19.66%] {2.6%}	3.26 (12.36) [20.51%] {0.9%}	2.26 (14.68) [35.04%] {4.3%}
Job hrs/week	0 (7.14) [9.41%] {0%}	2.69 (7.81) [50.42%] {0%}	1.72 (10.93) [26.49%] {0%}	1.84 (7.71) [29.06%] {0.9%}	2.26 (8.34) [47.86%] {2.7%}
Salary at the age of 30	-4928.57 (56071.43) [1.71%] {0%}	-4757.99 (26718.8) [17.09%] {0%}	-100 (21927.6) [16.23%] {0%}	22878.97 (37460.7) [16.24%] {0%}	12856.38 (40097.75) [16.24%] {0%}
No. of Observations	14	103	58	59	117

(.) mean absolute change in belief between the two surveys

[.] proportion of respondents for whom change in beliefs is ≤ 5 for binary outcomes; ≤ 5 for hrs/week; ≤ 5000 for inc

{.} Proportion of responses that are not a multiple of 5

a A major that the student was pursuing when first surveyed, but dropped at the time of the second survey

b An individual's least preferred major at the time of the second survey

c The second most preferred major for individuals without a second major

d The individual's second major

Table 8: The Nature of Change in Beliefs for Outcomes

Dependent Variable: Change in belief for:		Grad in 4 Years	Grad w/ GPA ≥ 3.5	Enjoy Courses	Course Hrs/Wk	Parents Approve	Find Job	Enjoy Work	Work Flexible	Job Hrs/Wk	Salary at 30
Constant	1.48 (1.13)	-5.32*** (2.05)	-4.11*** (1.50)	-5.53*** (1.24)	0.39 (1.51)	-0.92 (2.25)	-4.55*** (1.75)	2.05 (2.01)	2.13 (1.19)	14549*** (5227)	
Second Pursued Major	2.36* (1.42)	3.07 (2.50)	3.72 (2.44)	1.15 (1.41)	-0.19 (2.82)	-0.18 (3.13)	-1.37 (2.69)	1.82 (2.56)	0.097 (1.73)	7155 (12976)	
Second Preferred Major	-1.72 (1.47)	0.78 (2.78)	1.42 (3.16)	2.52* (1.16)	-5.51* (2.89)	-1.89 (2.84)	0.55 (3.38)	2.34 (3.13)	-1.03 (2.22)	-13982** (6797)	
Dropped Major	-8.48* (5.14)	6.84 (4.91)	-6.20 (4.34)	-0.31 (3.24)	-2.31 (3.44)	0.67 (2.74)	-2.01 (7.02)	1.98 (7.11)	-1.79 (3.23)	-20526 (27206)	
Least Preferred Major	-2.07 (2.23)	-1.23 (2.81)	-5.05* (2.72)	5.31*** (1.29)	-0.76 (2.58)	-1.24 (2.74)	-8.23*** (2.95)	-1.45 (2.87)	0.45 (1.39)	-19453*** (5565)	

Cluster standard errors in parentheses. * sig at 10%; ** sig at 5%; *** sig at 1%

Regressions include random effects. Each of these regressions has 341 observations with 117 groups (students).

The binary outcomes (all outcomes excluding coursework hrs/wk; job hrs/week; salary at 30) are on a 0-100 scale.

Table 9: Are individuals aware of changes in beliefs?

Perceived Change: Actual Change	Least Pref Major			Second Major			Current Major		
	Inc.	Unchg	Dec.	Inc.	Unchg	Dec.	Inc.	Unchg	Dec.
Graduate in 4 years									
Increase	3*	22	5	8	23	0	6	16	0
Unchanged	5	46	1	4	66	1	10	74	4
Decrease	2	20	9	0	9	6	0	2	1
Graduate with a GPA ≥ 3.5									
Increase	5	17	7	9	14	6	11	8	3
Unchanged	4	28	8	8	25	8	9	21	9
Decrease	1	29	15	4	29	14	9	20	23
Enjoy Coursework									
Increase	5	13	8	11	17	4	10	11	5
Unchanged	0	21	8	5	30	4	16	25	0
Decrease	8	38	13	7	20	18	11	25	10
Coursework hrs/week									
Increase	9	12	3	5	7	1	4	9	1
Unchanged	10	44	4	14	36	7	13	30	5
Decrease	7	20	5	11	27	8	20	21	10
Approval of Parents									
Increase	5	32	0	5	27	2	5	29	0
Unchanged	3	39	4	7	29	5	2	51	4
Decrease	4	20	7	5	34	3	3	15	4
Finding a job									
Increase	11	21	1	6	19	6	9	20	6
Unchanged	3	32	4	5	29	12	6	23	8
Decrease	2	31	9	4	25	11	7	21	13
Enjoying work at jobs									
Increase	3	12	9	9	21	3	12	8	5
Unchanged	1	19	4	7	22	5	15	26	5
Decrease	7	46	13	2	35	13	6	30	6
Reconciling work & family									
Increase	10	37	4	6	28	5	9	24	10
Unchanged	3	18	3	8	31	8	10	25	6
Decrease	2	33	4	4	23	4	4	12	13
Job hrs/week									
Increase	9	23	1	11	17	2	21	14	1
Unchanged	10	57	1	7	54	4	16	35	5
Decrease	1	8	4	3	13	5	6	14	1

*Each cell shows the number of respondents who fall in that category. Recall there are a total of 117 respondents.