Are there non-cognitive differences between first-generation college students and their peers that explain economic differences?

By Sharon Ma

This paper will examine why there are differences in key economic outcomes between first-generation and non-first-generation college students. Through both an analytical and experimental approach, this paper shows that first-gen college students are less patient and riskier than non-first-gen college students, on average, potentially serving to explain post-graduation income disparities.

I. Introduction

In order to explore generational differences, this paper will explore the differences between non-first-generation college students and first-generation college students, in terms of income and employment. It will use an analytical method, which uses the SHED data to see whether the two groups, on average, tend to enter different careers, industries, and types of companies; which may lead to income disparities. It will also use experimental methods, including priming from psychology to see whether the two groups perform differently -- due to differences in patience, risk aversion, confidence, and competitiveness -- after they are made conscious of their first-gen (or non-first-gen) and socioeconomic statuses. The results of the experimental approach indicate that first-gen college students are riskier and less patient than non-first-gen students, on average, which contradicts some of the existing literature (Trejo, 2016).

To begin looking into this topic, it is important to understand that this trend exists in the first place: There has been a lot of research on the effect of parental education on their children's education. Most studies agree that there is a positive relationship between the two: the higher the educational attainment of a parent, the higher the educational attainment of the child (Emmons et al, 2019). The effect may be even stronger when both parents, not just one of them, have a high education level. These positive results may be due to the fact that parents who have gone to college themselves know the college admissions process, (Choy, 2001) as well as how to navigate through college life (Emmons et al, 2019). In other words, they know what Advanced Placement (AP) or Scholastic Aptitude Tests (SAT) exams their children should take, as well as what they should do and who they should talk to once they are in college. These positive results may also be due to college-educated parents pushing their children to become educated as well, due to them being more aware than anyone else the long-term benefits of obtaining a Bachelor's degree. This is good news for a lot of students with educated parents because they are likely to attend a selective, 4-year university (Emmons et al, 2019). However, what about first-gen college students? Are they "destined" for failure, just because their parents did not attend college?

Many people deem college to be "the great equalizer" because, as mentioned before, they believe that a bachelor's degree yields long-term benefits, such as higher incomes and wealth, regardless of the students' or parents' educational backgrounds. In other words, intuitively, obtaining a bachelor's degree would put first-gen college students on the same footing as their peers, allowing them to obtain similar incomes. However, a study by the Pew Research Center says otherwise, finding that there is still a gap between first-gen and non-first-gen college graduates in their economic outcomes. (Pew Research Center, 2021) If anything, college made the gap between first-gen and non-first-gen students even greater because when the two groups' highest educational attainments were junior high or high school, they were earning similar incomes. It is only after graduating from college that the latter makes significantly more than the former: $135,800, compared to $99,600. (Figure 1)

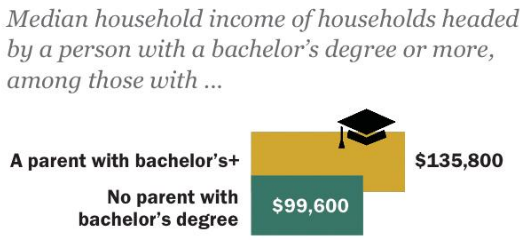


Figure 1. Relationship Between Parental Highest Educational Attainment and Children Income. This graph by the Pew Research Center (2020) shows that the median three-person household income of college graduates ages 22 to 59 from the 2019 Survey of Household Economics and Decisionmaking whose parent(s) have at least a Bachelor's degree was $135,800 while those whose parents did not have a Bachelor's degree earned $99,600. (Source: Pew Research Center, 2020)

So if we've already ruled out having a college degree as being "the great equalizer", then why does this large difference still exist? To explore this question, this paper uses an analytical and experimental approach. Based on the analytical approach, differences in the career and industry choices between first- and non-first-gen college graduates can be seen. Based on the experimental approach, differences in their patience, risk aversion, confidence, and competition levels can be seen. The combination of these two analyses show that first-gen college students are less patient and riskier than non-first-gen college students, on average, potentially serving to explain post-graduation income disparities.

II. Literature Review

Previous research on first-generation college students finds that there exists a "parent premium," which is the boost in income from having a parent who has at least a Bachelor's degree (Pew Research Center, 2021, p. 10). This same study finds that college graduates, ages 22 to 59 in the 2019 Survey of Household Economics and Decisionmaking (SHED), who have a college-educated parent (non-first-gen college graduates) had a three-person household income of approximately $135,800 while those who did not have a college-educated parent (first-gen) earned $99,600. This income gap could be even bigger as non-first-gens are more likely to further their education (43% versus 35%), which means that they are likely to earn even more (up to $162,000). Richard Fry (2021), the author of this study, summarizes many economic differences between first-gen and non-first-gen college students by subsetting the sample into the two groups and then determining the averages for each variable. Taking an analytical approach, my paper looks into the updated version of the SHED dataset and digs deeper into the many variables to see whether there are other differences that are not already explained in the Pew Research Center study that could help explain the economic differences that Fry finds between first-gens and non-first-gens.

Taking an experimental approach, this paper draws from Benjamin et al (2020), who further draws from "self-categorization theory" by William James (1890, p. 221), which is the idea that researchers can use proxies to make participants more aware of their social identities, which drives them to act in ways that coincide with the norms associated with that social identity (Benjamin et al, 2010). They followed this idea by indirectly prompting the treatment group to be more aware of their social identity, and then investigated the differences in patience and risk aversion between the treatment and control groups across racial and gender groups. Their results supported James' self-categorization theory by seeing that primed Asian Americans followed their norm for patience and primed females followed their norm for risk aversion. These cognitive behaviors may explain their differences in negative economic outcomes, such as educational attainment, income, savings, and wealth. This paper uses the mechanism of priming by seeing whether priming first-gen college students to be aware of their status, for example, would make them subconsciously answer the patience, risk aversion, confidence, and competitiveness games differently.

When it comes to the non-cognitive behaviors of patience, risk aversion, confidence, and competitiveness; this idea was inspired by the paper above, as well as one by Algan et al (2022). The latter explored whether the development of non-cognitive skills in early childhood is linked to long-term social and economic outcomes. Algan et al (2022) did this by analyzing the Montreal Longitudinal Experimental Study (MLES) (1984), where 1,037 boys from 53 low socioeconomic status schools were randomly assigned the treatment, which was an invitation to participate in a two-year training program that assists with social skills and self-control. One would think that these factors would not contribute to higher income later on, but the results of this paper says otherwise: By connecting past MLES (1984) data with recent income and tax data, Algan et al (2022) find that the treatment resulted in increased grades, years of education, marriage rates, employment rates, and income; amongst others. Although this article by Algan et al (2022) isn't about first-gen students, it is still useful because it shows that non-cognitive behaviors are connected to economic outcomes, and so the potential differing non-cognitive behaviors between first-gens and non-first-gens may be the reason for their differing economic outcomes later on.

When it comes to the patience, risk, confidence, and competitiveness games; the core methods being used come from the papers by Harrison et al (2002) and Boneva et al (2021). The first paper by Harrison et al (2002) uses a price list to gauge patience. A price list usually follows the format of $100 today or $100+x tomorrow. If the subject prefers $100 today, then they are impatient, and their discount rate is defined as x% per day. My paper uses this same price list method but with different numbers and scenarios. On the other hand, the second paper by Boneva et al (2021) asks participants to self-report their competitiveness using a linear scale, finding differences in earning expectations tied to differences in competitiveness and gender. My paper took these questions into consideration as well.

With these experimental methodologies, non-cognitive behaviors, and experimental games and questions in mind; this paper hypothesizes that first-gen college students are less patient (more likely to want to start working right away, rather than investing the time and money into attending university) and more risk averse (for example, they would be less willing to take out a large student loan to go to university). These hypotheses are backed by the article by Sam Trejo (2016), which uses data from the National Longitudinal Survey of Youth (1997) to find that first-gen college students are more risk averse since they elect into majors that are correlated with more job security, higher expected wages, and clearly defined paths (Trejo, 2016). My experimental design includes games and questions that measure major choices, job security, and expected wages; which would allow me to see whether or not first-gen college students truly elect into different majors and careers than their peers.

III. SHED Analysis

A. Data

In an analytical attempt to find differences between first-generation and non-first-generation college graduates that were not originally found in the Pew Research Center (2021) article that could potentially explain their differing economic outcomes, my paper used the updated 2020 Survey of Household and Economic Decisionmaking (SHED), conducted by the Federal Research Board, to better understand the average first-gen versus non-first-gen. As its name suggests, this survey attempted to measure the economic well-being of 11,648 individuals through 372 questions, which include but are not limited to: highest educational attainment, income, and employment variables (Federal Reserve, 2020). The following specific variables were analyzed:

- *Mother education* - the highest level of education in which the individual's mother has completed,
- *Father education* - the highest level of education in which the individual's father has completed,
- *Individual education* - the highest level of education in which the individual has completed,
- *Household head* - whether or not the individual is the head of the household in which they reside,
- *Household income* - the household income,
- *Employment status* - the individual's current employment status,
- *Industry Type* - the individual's main or current industry,
- *Employer Type* - the individual's type of employer.

B. Confirmation: Regression, Graduation Rates, and Income

As a quick confirmation of the findings by the Pew Research Center (2021), *mother education* and *father education* were regressed onto *individual education*:

$$individual\ education_i = \beta_0 + \beta_1 *\ mother\ education_i + \beta_2 *\ father\ education_i + \varepsilon_i,$$

The following table is the regression results, all of which are statistically significant at the 1% level:

Table 1: Regress individual on mother and father education

	Dependent variable:
	Individual Education
Mother education	0.191***
	(0.010)
Father education	0.218***
	(0.010)
Constant	2.100***
	(0.023)
Observations	10,392
R^2	0.183
Adjusted R^2	0.183
Residual Std. Error	0.826 (df = 10389)
F Statistic	1,165.957*** (df = 2; 10389)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2: Regress *individual education* on *mother education* and *father education*. At the 0.01 significance level, a one unit increase in the mother's highest educational attainment increases the individual's highest educational attainment by 0.191. This effect is even greater at 0.218 from the father's side.

This table shows that holding the father's highest educational attainment constant, a one unit increase in the mother's highest level of education increases the child's highest level of education by approximately 0.191, on average. Meanwhile, holding the mother's highest educational attainment constant, a one unit increase in the father's highest level of education increases the child's level of education by approximately 0.218, on average. The father's highest educational attainment seems to have a slightly larger effect on the child's than the mother's, which could be due to the patriarchal nature of some households. These statistically significant results show that a parent's highest educational attainment does indeed play a positive role in the child's highest educational attainment. A parent obtaining their bachelor's degree can very much bridge the gap between whether or not the child obtains their bachelor's degree, as well.

Another way to determine whether or not there is a relationship between parental education and child's education is by looking at the graduation rates themselves:

Number of College Graduate Parents vs Graduation Rates	
Students whose	Graduation Rate
Parents both graduated from college	79.103%
Father or mother graduated from college	69.05%
No parents graduated from college	32.006%

Figure 3: Number of Parents who Graduated from College versus Graduation Rate. As the number of parents who have graduated from college increases from 0 to 2, the graduation rates increase from 36% to 79%.

As you can see, approximately 79.103% of non-first-gen students graduate from college and obtain their bachelor's degrees, 69.05% of students whose fathers or mothers have graduated from college also graduate from college, and only 32.006% of first-gen students graduate from college. Although this data is not conditional on going to college in the first place, the difference is still huge (47.097%), and it confirms that there is indeed a correlation between a parent's highest level of education and their child's.

Last but not least, the average adjusted 3-person household incomes of first-gen and non-first-gen college graduates were analyzed. This resulted in approximately $97,000 for first-gens while it was higher at $109,000 for non-first-gens. These results confirm the differences that were found by the Pew Research Center (2021).

C. Extension: Employment

In an analytical attempt to extend the Pew Research Center's (2021) findings, *employment rate, industry type*, and *employer type* were analyzed. It turns out that only 67% of first-generation college graduates were employed while 86% of non-first-generation college graduates were employed. This difference could be due to a myriad of factors, such as non-first-gens having more resources. Their parents could have a better network and/or be able to answer questions regarding their careers, as explained by Ana Hernandez Kert in her Federal Reserve Bank of St. Louis article (2021).

It also appears that when it comes to listing their main or current industry, a high proportion of first-gen college graduates in the 2020 SHED (approximately 33%) listed "NA" when answering the survey. There could be many reasons for this choice. They could be unemployed. Or they could be moving between industries, and thus, haven't grounded themselves into any specific one yet. They may also be in an uncommon and unidentifiable industry. Unfortunately, unemployment, fewer years of experience in an industry, and obscure industries are usually associated with lower income.

Regardless of what the reason may be, there is still a clear difference in the industries that non-first-gen college graduates enter versus first-gen college graduates. A clear majority is in one category: approximately 21% in the "Professional, Scientific, Technical, and Businesses Services" industry. This industry has an average annual salary of $101,287 (Bureau of Labor Statistics, 2021) while the general average annual salary in the US is $51,916.27 (Bureau of Labor Statistics, 2021). As you can see, this industry pays more than the average industry, and the largest proportion of non-first-generation college graduates happen to be in this industry. Meanwhile, only 13% of first-generation college graduates are in the "Professional, Scientific, Technical, and Businesses Services" industry, indicating a gap between first-generation and non-first-generation students, even after graduation.

Lastly, there is also a clear difference in the types of employers that non-first-gen college graduates work for, compared to their first-gen counterparts. As in the earlier statistic, a good chunk (33%) of first-gen college graduates tend to be unemployed. This is reflected in the 41% of first-gen college graduates who either did not have an employer or did not know their type of employer. An example for the latter may be that the individual has a non-traditional job, such as being a contracted worker. However, as before, non-first-gen college graduates didn't have this problem -- a large proportion of them (approximately 46%) work at private-for-profit companies, which again tends to pay more than any other type of employer. On the other hand, only 32% of first-generation college graduates work at private-for-profit companies, again, indicating a gap.

As one can see, the average non-first-generation college graduate tends to be employed at a private-for-profit company that is in the "Professional, Scientific, Technical, and Businesses Services" industry while the average first-generation college graduate does not, which may explain why the average income for non-first-generation college graduates is higher than the latter.

IV. Experiment Trial #1[1]

A. Experiment Design

This experiment was conducted on Qualtrics, which is a web-based surveying and report-generating platform that allows the use of randomized blocks to assign individuals into the treatment or control groups. Data was collected from 279 University of Michigan undergraduate students, and participation was requested via emails that were sent out to first-generation student organizations on campus and miscellaneous undergraduate majors listservs, such as economics, data science, architecture, psychology, mathematics, and physics. A guaranteed gift card upon completion was explicitly stated in the emails. It is important to note, however, that the budget was limited to only $200 from the Economics Department, and thus, each student received approximately only $0.88. This means that the results could have been even more significant, if

[1] This section was pulled from the "Final Paper" from the University of Michigan course "Econ 408: Experiments in Econ", written with Sophie Shao.

there was a larger budget, since there would have been higher incentives for more students to complete the experiment, generally and more accurately.

Approximately half of the participants (111) received the priming treatment, in which they were asked questions that made their first-generation student status, or lack thereof, salient. These priming questions were the following:

- What is your mother's highest education level?
- What is your father's highest education level?
- What proportion of students at the University of Michigan do you believe are first-generation college students, meaning that they are the first in their immediate family to attend college?
- How much support did you get from your family in applying for/going to college?
- Do you think college is more or less challenging for first generation college students?

The other half (110), on the other hand, received the control condition, in which they were asked miscellaneous questions unrelated to their non-first-gen or first-gen status, and thus should not have made this status salient. This means that when a participant is asked these questions, their first thought should not be about their first-gen status, due to the lack of connection between the two topics. These control questions were the following:

- Which of the following streaming services do you use?
- Which campus library do you utilize the most?
- How many course credits are you taking this semester?

Next, all participants, regardless of whether they received the treatment or control, were asked to perform a few games that measure patience and risk aversion. Afterwards, they were asked to complete a post-survey that explicitly asks about their first-gen student status and gauges their attitudes towards higher education. The details behind these games, such as what they entail and the reasons behind their features, will be elaborated on in Sections C to F. In addition, the post-survey was purposely placed at the very end to ensure that the questions here would not accidentally prime the control group as well. For the full version of this survey, see Appendix A.

With prior knowledge, related literature, and experimental design; we hypothesized that first-generation college students would be less patient and more risk-averse than non-first-generation college students since they would prefer to work right away, rather than pursue higher education (less patient), and choose majors that are usually associated with defined careers and higher job security.

B. Data

279 University of Michigan undergraduate students participated in the experiment, but due to missing answers, the data was cleaned to just 221, in which 44 of them were first-generation and 177 of them were non-first-generation. Approximately half of these participants were randomly assigned into the treatment group and the other half into the control, resulting in 23 treated first-gens, 21 control first-gens, 88 treated non-first-gens, and 89 control non-first-gens. A chart that summarizes this breakdown can be seen below.

Breakdown of Sample (Counts)		
	First-Gen	Non-First-Gen
Treatment	23	88
Control	21	89
Total	44	177

Figure 4: Breakdown of Sample. This chart shows the number of first-gens versus non-first-gens in the treatment versus control groups in counts.

The two groups, at the University of Michigan at the very least, are very similar. They are both predominantly female, with 59% of first-gens identifying as female, a proportion that is very similar to the 61% for non-first-gens. This statistic is not surprising because according to a study by the Federal Reserve Bank of St. Louis, more women are enrolled in college than men these days (Leukhina and Smaldone, 2022).

They are also both predominantly White and Asian, with 45% of first-gens and 49% of non-first-gens being White and 43% of first-gens and 44% of non-first-gens being Asian. Again, this statistic is not surprising because it is consistent with enrollment statistics at the University of Michigan. That being said, because the University of Michigan is a skewed representation of the general population, the results of this experiment may not accurately represent all first-gens and non-first-gens in the U.S.

Lastly, both groups are predominantly upperclassmen, with 41% of both first-gens and non-first-gens in the sample being seniors. Ideally, the sample should have been mostly underclassmen, but because underclassmen at the University of Michigan are less likely to be on majors listservs and less likely to know how to participate in school experiments, it was harder to reach them.

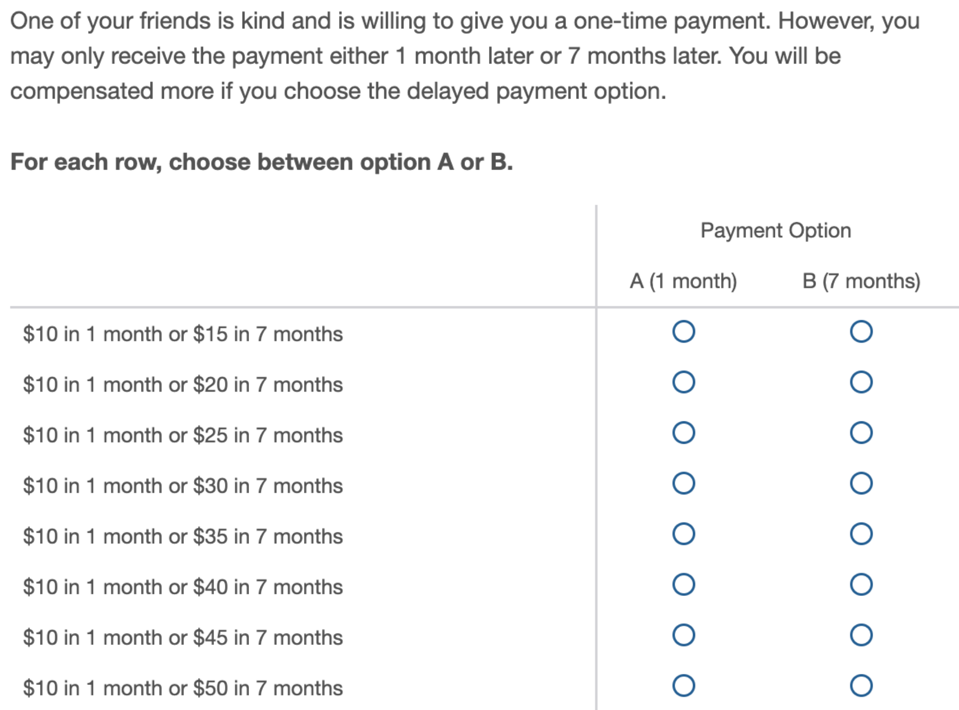
However, the fact that the compositions of the two groups are so similar to each other in the sample is a good sign. This means that the two groups are comparable and that the results that are shown in this paper are mostly due to the participants' generation status and not their gender, race, or school year. A chart that summarizes the descriptive statistics of the sample can be seen below.

Descriptive Statistics of Sample (Percentage)		
	First-Gen (44)	Non-First-Gen (177)
Female	59%	61%
Male	39%	37%
Other	2%	2%
White	45%	49%
Asian	43%	44%
Black	2%	1%
Other	9%	6%
Freshman	16%	6%
Sophomore	18%	15%
Junior	23%	38%
Senior	41%	41%
Other	2%	1%

Figure 5: Descriptive Statistics of Sample. This chart breaks the first-gen and non-first-gen groups down by gender (female, male, other), race (White, Asian, Black, other), and school year (freshman, sophomore, junior, senior, other) in percentages.

C. Patience Game

The patience game that all participants had to take, regardless of whether they were in the treatment or control group, is the following:



One of your friends is kind and is willing to give you a one-time payment. However, you may only receive the payment either 1 month later or 7 months later. You will be compensated more if you choose the delayed payment option.

For each row, choose between option A or B.

	Payment Option	
	A (1 month)	B (7 months)
$10 in 1 month or $15 in 7 months	○	○
$10 in 1 month or $20 in 7 months	○	○
$10 in 1 month or $25 in 7 months	○	○
$10 in 1 month or $30 in 7 months	○	○
$10 in 1 month or $35 in 7 months	○	○
$10 in 1 month or $40 in 7 months	○	○
$10 in 1 month or $45 in 7 months	○	○
$10 in 1 month or $50 in 7 months	○	○

Figure 6: Patience Game. The participants are asked whether they want to receive $10 in 1 month or $10+*x* in 7 months. This *x* increases as they go down the list.

This game was included in the experiment to measure the participants' patience levels. If the participant is extremely impatient, they would take the $10 in 1 month, no matter the value being offered in 7 months. If the participant is extremely patient, they would take the $10+*x* in 7 months. The point at which a participant switches from A to B (from choosing $10 in 1 month to $10+*x* in 7 months) measures their patience level using the discount rate *x*% per day. The sooner the participant makes the switch, the more patient they are.

One way in which this game was analyzed was by comparing the proportion of first-generation college students who would take the $10 in 1 month (a.k.a. be impatient) to that of non-first-generations. The same was done within groups as well, between primed versus controlled. This resulted in the two graphs below: one consisting of just first-gen students (left) and the other, non-first-gen (right). Within each graph, they were further divided into whether they were primed (blue) or in the control group (red). In the graph below, the x-axes are the 7-month values, and the y-axes are the proportions of participants who chose $10 in 1 month.

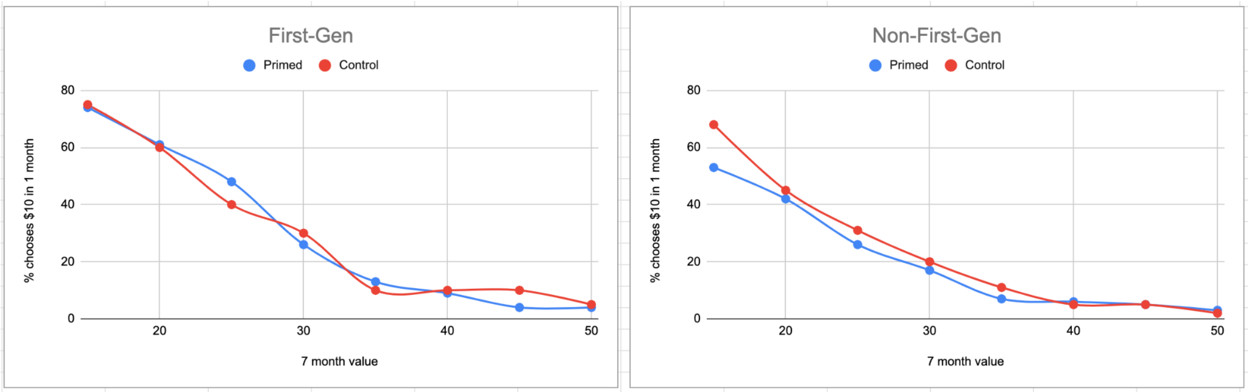


Figure 7: Proportion of Primed and Control First- and Non-First-Gen Who Choose $10 in 1 Month. The graph on the left consists of the first-gens in the sample while the graph on the right consists of the non-first-gens. Within each group, the blue lines represent those who received the priming treatment while the red represents those who received the control.

As shown in the first-gen graph (left), there doesn't seem to be an obvious effect of the priming treatment on them, as there is little to no gap between the two lines. There does, however, seem to be a difference between the first-gen (left) and non-first-gen (right) graphs, so a Welch Two-Sample t-test was conducted to determine whether or not this difference is statistically significant. The reported patience level for first-gens was approximately 4.64 while that for non-first-gens was 4.311. The p-value for this test is 0.57674, which is not less than 0.05. But it is good that we cannot reject the null hypothesis because it suggests that our hypothesis that first-gens are less patient than non-first-gens may still stand.

Another way in which this game was analyzed was by running the following regression. The fixed effects here would determine whether there is a difference in being in the primed group versus control.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \text{, where}$$

$y = patience$ (a score determined by the Patience Game),

$x_1 = non\ first\ gen$ (1 if non-first-gen, 0 if first-gen), and

$x_2 = primed$ (1 if primed, 0 if controlled).

The coefficient of interest is β_3, which can be interpreted as the difference in the patience levels between primed and controlled first-gens, capturing the effect of priming.

This regression resulted in a statistically significant mean patience score of 5.286 for the controlled first-gens. When they are primed, their patience score increases by 0.062 -- to 5.349, though this is not statistically significant. On the other hand, when the participant is in the non-first-gen group, their mean patience score decreases by 0.218 -- to 5.068, though this is also not statistically significant. Lastly, when a participant is both primed and non-first-gen, their mean patience score is 5.489, though unfortunately, this is also not statistically significant. A summary of these regression results are shown below.

Table 1: Patience Game

	Dependent variable:
	Patience
Primed	0.062
	(0.903)
Non-first-gen	−0.218
	(0.726)
Primed and Non-first-gen	0.359
	(1.008)
Constant	5.286***
	(0.653)
Observations	221
R^2	0.004
Adjusted R^2	−0.010
Residual Std. Error	2.991 (df = 217)
F Statistic	0.296 (df = 3; 217)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 8: Patience Game Regression Results. *Patience* is regressed on *primed* and *generation.*

D. *Risk Game #1*

One of the risk games that all participants had to take is the following:

Congrats, you've made it to the final round in a popular game show! You may go home with $10,000 guaranteed if you choose, but if you participate in the challenge round, you may have the opportunity to win even more: $20,000! For each of the following chances, would you accept or decline the challenge round?

	Accept/Decline Challenge Round	
	Accept	Decline
Accept 10% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 20% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 30% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 40% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 50% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 60% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 70% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 80% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 90% chance of $20,000 or Decline with $10,000 guaranteed	○	○

Figure 9: Risk Game #1. All participants are placed in a scenario in which they have to choose between a chance of $20,000 or $10,000 guaranteed. This chance increases as they go down the list.

This game was included to measure the participants' riskiness levels. If the participant is extremely risk averse, they would take the $10,000 guaranteed, no matter what. If the participant is extremely risky, they would take the chance to win $20,000, no matter what. The point at which the participant switches from $10,000 guaranteed to the chance to win $20,000 represents their risk score for this game. The sooner the participant makes the switch, the riskier they are.

One way in which this game was analyzed was by conducting a Two-Sample Test for Equality. The mean first-gen risk score was 4.090909 while that for non-first-gen was 4.361582. This difference is statistically significant at the 10% level since the p-value is 0.09376, which suggests that there is a 90% chance that first-gens have a lower risk score of 4.090909, compared to the 4.361582 for non-first-gens, supporting our hypothesis that first-gens are less risky than non-first-gens.

A Welch Two Sample t-test was then conducted to see whether those who were primed were more risk averse, conditional on being first-gen. The mean risk score for primed first-gens was 3.714 while that for controlled first-gens was 4.435. This difference, unfortunately, is not statistically significant, due to the p-value being 0.362. But if it were, it would have supported our hypothesis that priming first-gens to be aware of their first-gen status would make them even more risk averse.

Another way in which this game was analyzed was by running the following regression:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = risk1$ (a score determined by Risk Game #1),

$x_1 = not\ first\ gen$ (1 if non-first-gen, 0 if first-gen), and

$x_2 = treatment$ (1 if primed, 0 if controlled).

Although the results are not statistically significant, a summary of them are shown in Figure 12.

E. Risk Game #2

Another risk game that all participants had to take is the following:

You are a sales associate for a small boutique, and you need to decide which days of the week that you have to go in for work. Because your job is commissions based, however, there is no guarantee that you will get the full amount. The following chart shows the potential payoffs from each shift schedule. Which of the following shift schedules would you choose?

- ○ ½ of $80, ½ of $12
- ○ ½ of $70, ½ of $22
- ○ ½ of $62, ½ of $26
- ○ ½ of $54, ½ of $30
- ○ ½ of $46, ½ of $34
- ○ ½ of $38, ½ of $38

Figure 10: Risk Game #2. All participants are placed in a scenario in which they had to choose which days of the week to work. Certain days have high payoffs while other days have low.

In this game, participants were placed into a scenario in which they were sales associates who had to choose among 6 different lotteries associated with different shift schedules. In each lottery, the participant has a 50% chance of earning a higher amount and a 50% chance of a lower amount. As the participants go down the list, the higher and lower amounts get closer to each other, generating a guarantee of some modern amount. If the participant is extremely risk averse, they would choose the last choice (½ of $38, ½ of $38). If the participant is extremely risky, they would choose the first choice (½ of $80, ½ of $12). In other words, risk score decreases as participants go down the list.

Similar to Risk Game #1, a Two-Sample Test for Equality was conducted. The mean first-gen risk score was 3.07 while that for non-first-gen was 3.08. This difference, unfortunately, is not statistically significant, due to the p-value being 0.169.

Similar to before as well, the following regression was done:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = risk2$ (a score determined by Risk Game #2),

$x_1 = not\ first\ gen$ (1 if non-first-gen, 0 if non-first-gen), and

$x_2 = treatment$ (1 if primed, 0 if controlled).

Fortunately, all of these results are statistically significant, as seen in Figure 12. Here, controlled first-gens have a mean risk score of 2.476. When primed, their risk score increases by 1.133 -- to 3.609. Meanwhile, controlled non-first-gens have a mean risk score of 3.2250, and primed non-first-gens have a mean risk score of 1.811. These results support our hypothesis that first-gens are less risky than non-first-gens (2.476<3.2250). Interestingly, however, making non-first-gens aware of their identity makes them a lot less risky. Perhaps asking non-first-gens questions related to first-gens instead makes them put themselves into the shoes of first gens, making their shoes, making the non-first-gens act in less risky ways.

F. Risk Game #3

The last risk game is the following:

You received multiple job offers. As you go down your list of job salaries, the salaries significantly increase, from as little as $40,000 to as much as $70,000. However, as the salary increases, the probability that you get laid off within the first few months (and thus, do not get paid the full salary) increases as well. You can only accept one job offer at this time. Which of the following job offers would you choose?

- Salary of $40,000 with 0% probability of being laid-off
- Salary of $75,000 with 10% probability of being laid-off
- Salary of $100,000 with 20% probability of being laid-off
- Salary of $200,000 with 30% probability of being laid-off

Figure 11: Risk Game #3. All participants are placed into a scenario in which they have to choose among different job offers. Certain job offers have a lower salary, but lower chance of being laid off.

In this game, participants were placed into a scenario in which they had to choose among different job offers. As they go down the list of job offers, the salaries significantly increase, but the probability that they get laid off within the first few months (and thus, do not get paid the full salary) increases as well. If participants were extremely risk averse, they would choose the first job offer ($40,000 with 0% probability of being laid off). If the participants were extremely risky, they would choose the last job offer ($100,000 with 30% probability of being laid-off). Unfortunately, there was a typo in this question, but it is later corrected in Section V: Experiment Trial #2.

Similar to before, the following regression was done:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = risk3$ (a score determined by Risk Game #3),

$x_1 = not\ first\ gen$ (1 if non-first-gen, 0 if first-gen), and

$x_2 = treatment$ (1 if primed, 0 if controlled).

However, most likely due to the typo, there were no statistically significant differences, as most people were inclined to choose “$200,000 with a 30% probability of being laid off”. A summary of these results are still shown in Figure 12.

G. Aggregate Risk Score

Although some games on their own did not produce statistically significant results, when they were aggregated, the coefficients all became significant. This may be due to differences being small and underpowered on their own, but once they are aggregated, these differences become big enough to be considered significant.

Risk Games Regression

	Dependent variable:			
	Risk Game 1	Risk Game 2	Risk Game 3	Aggregate Risk Score
	(1)	(2)	(3)	(4)
Treatment	0.720	1.133**	0.720	1.905***
	(0.503)	(0.518)	(0.503)	(0.727)
Not First-Generation	0.623	0.749*	0.623	1.826***
	(0.404)	(0.416)	(0.404)	(0.584)
Treatment and Not First-Generation	-0.671	-1.414**	-0.671	-2.201***
	(0.562)	(0.579)	(0.562)	(0.812)
First-Generation (Constant)	3.714***	2.476***	3.714***	9.095***
	(0.364)	(0.374)	(0.364)	(0.526)
Observations	221	221	221	221
R^2	0.014	0.027	0.014	0.046
Adjusted R^2	0.0001	0.013	0.0001	0.032
Residual Std. Error (df = 217)	1.667	1.716	1.667	2.409
F Statistic (df = 3; 217)	1.007	1.992	1.007	3.457**
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 12: Aggregated Risk Chart. This chart shows the regression results from Risk Games 1-3, as well as the Aggregate Risk Score.

Now, the mean aggregate risk score for controlled first-gens was 9.095. When primed, their risk score increases by 1.905 -- to 11. Meanwhile, the mean aggregate risk score was 10.92 for controlled non-first-gens and 8.72 for primed non-first-gen. These results suggest that non-first-gens are riskier than first-gens, supporting our hypothesis. Interestingly, priming non-first-gens to be salient of their identity actually decreased their risk score, a result that was shown in Risk Game #2. Again, perhaps asking non-first-gens questions related to first-gens actually made them aware of their privilege, convincing them to momentarily be in the shoes of a first-gen and making them behave in a more risk averse manner.

H. Post-Survey

After completing the games, the participants then completed a series of questions that gauges their expected years for degree completion, expected and desired earnings post-graduation, post-graduation plans, and perception of college as a positive or negative investment.

The most interesting result from these questions came from the following regression:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = sentiment$ (a score determined by the post-survey),

$x_1 = not\ first\ gen$ (1 if non-first-gen, 0 if first-gen), and

$x_2 = treatment$ (1 if primed, 0 if controlled),

which produced the following results:

Sentiment Score Regression

	Dependent variable:
	Sentiment Score
Not First-Generation	-0.005
	(0.151)
Treatment	-0.325*
	(0.188)
Treatment x Not First-Generation	0.165
	(0.210)
First-Generation (Constant)	3.095***
	(0.136)
Observations	221
R^2	0.029
Adjusted R^2	0.016
Residual Std. Error	0.623 (df = 217)
F Statistic	2.182* (df = 3; 217)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 13: Regress *Sentiment* on *Primed* and *Generation*. At the 10% significance level, the priming treatment decreases sentiment scores by 0.325.

Here, the mean sentiment score for controlled first-generation college students is 3.095. When primed, this score decreases by 0.325 -- to 2.77, which supports our hypothesis because this suggests that making first-gens aware of their identity leads them to abide by their social expectations and have a lower perception of college. This change is statistically significant.

Meanwhile, the mean sentiment score is 3.09 for controlled non-first-gens and 2.765 for primed non-first-gens. Unfortunately, these numbers are very similar to the others and are not statistically significant.

Other interesting results came from the questions regarding expected and desired earnings, grouped by generation, as seen in the following boxplots:

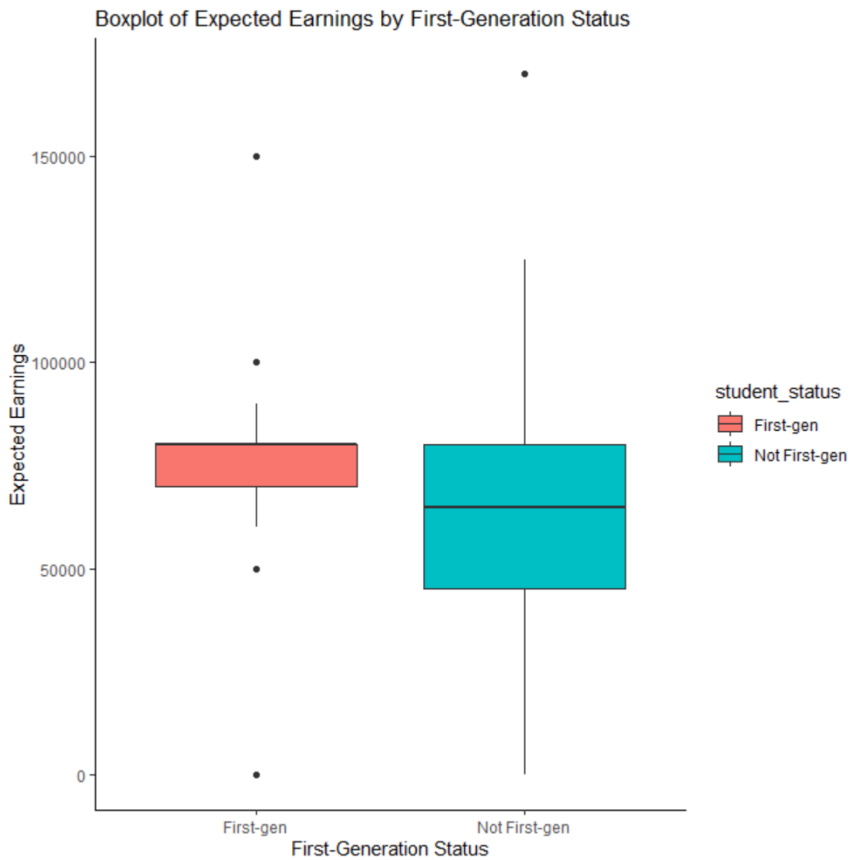


Figure 14: Expected Earnings by Generation. Although the median expected earnings for first-gens are higher, its distribution is smaller and consolidated, suggesting little to no variety in their major and career choices.

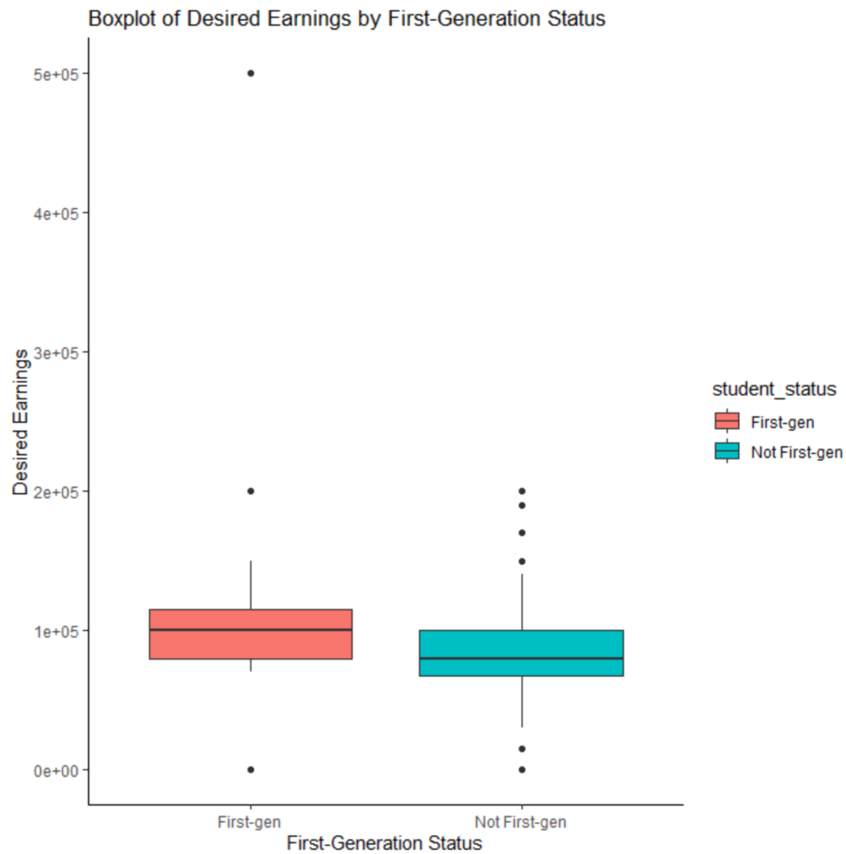


Figure 15: Desired Earnings by Generation. The medians and distributions are very similar.

The first boxplot shows that the median expected earnings for first-gens are actually higher than non-first-gens. One might think that this result contradicts existing literature that first-gens earn less on average (Pew Research Center, 2021), but it is important to remember that the sample of first-gens comes from the University of Michigan - Ann Arbor, which is a competitive school in which its average student has higher grit, intelligence, work ethic, etc. than the national average. This result, however, is congruent with literature that argues that first-gens choose majors that have defined career paths with higher earnings and job security (Trejo, 2016). This conclusion can be drawn from the first-gen distribution being smaller and more consolidated, suggesting little to no variety in major and career choices. The second boxplot reinforces the first, suggesting that careers aside, these University of Michigan - Ann Arbor students' desired earnings are the same. Again, these numbers may be influenced by the readily available information on University of Michigan - Ann Arbor average salaries post-graduation.

V. Experiment Trial #2

A. Experiment Design

Trial #1 produced promising results, but in order to test whether they are applicable in a real world setting, a robustness check was done through Trial #2. This trial uses the same concepts and analyses as the first, except many parts (e.g. typos) are fixed and the scenarios are now different so that participants who may have participated in both trials cannot guess the intent behind the study, if they happened to have figured it out after the first. This second trial also has a larger budget of $500, rather than just $200, so participants now have a higher incentive to take the time to think through their choices. The flip side to this, however, is that while the participants have a higher incentive to answer the questions more accurately, there is now also a higher chance of participants completing the survey solely for monetary reasons. By the nature of how many of the survey questions are related to money, it is important to keep in mind that having a sample that has these monetary-minded individuals may skew the results. In an ideal world, everyone would be exactly the same, including their likelihood of participating when there are monetary incentives, except for one thing: their generation status.

In this non-ideal world, participation was elicited via emails sent to undergraduate majors listservs -- economics, data science, architecture, psychology, math, and physics -- and first-generation student organizations and posted in university Reddits and Discord Servers. This resulted in 151 undergraduates from various campuses, but mainly University of Michigan - Ann Arbor. Some of the other campuses include University of Michigan - Dearborn, University of Wisconsin - Madison, and Boston University. These universities should be similar in their cultures and compositions. Ideally, there would've been more participants, but it was tough to do so, due to universities being on spring break at the time. As before, this experiment was conducted on Qualtrics. See Section IV: Experiment Trial #1 for more details.

Regardless, approximately half of the participants (60) received the priming treatment. The difference between the priming treatment this time versus last time is that instead of just including questions that make first-gen status salient, there are now also questions that make their socioeconomic status salient. Perhaps it's not just their first-gen status but also their low socioeconomic status, for example, that make participants behave a certain way.

I now hypothesize that the first-gen low-income students would be less patient, more risk averse, less confident, and less competitive. Perhaps they don't feel as adequately prepared, compared to their peers, academically and professionally, because of their limited financial resources. Having these socioeconomic status variables also helps subset the first-gen group even further since there is still a lot of heterogeneity, even within the first-gen group itself. For example, there are first-gen students who are low-income, which is what their status is usually associated with, but there are also first-gen students who come from rich backgrounds. These priming questions were the following:

- What is your mother's education level?
- What is your father's education level?
- How much support did you get from your family in applying for/going to college?
- How much pressure did you receive from your family to attend college?
- Approximately how much is your total household income? Round to the nearest thousand.
- How often does your family worry about paying the bills?

- Did you have to take out student loans?
- Approximately how much do you have to pay in tuition out of pocket per semester?
- Approximately how much of your tuition does your need-based financial aid cover?
- How many months after graduation do you believe is an acceptable time to start working full-time?
- If you are a first-generation college student, do you think navigating through college life is more or less challenging, compared to your peers?
- What proportion of University of Michigan students do you believe are first-generation students?

The other half (69 participants), on the other hand, received the control condition, in which they were asked miscellaneous questions unrelated to either their first-gen nor socioeconomic statuses, and thus, should not have made either of their statuses salient. These control questions were the following:

- What proportion of University of Michigan students do you believe have a dining plan of some sort?
- Which of the following streaming services do you use?
- Approximately how many novels have you read in 2022?
- Which of the following fast-food restaurants do you have a positive opinion of?
- Which of the following retailers do you have a positive opinion of?
- Which operating software does your phone(s) use?

As before, all participants, regardless of whether they were primed or controlled, completed the same patience, risk, confidence, and competitiveness games and post-survey that explicitly asks about their first-gen status and gauges their attitudes and expectations towards college and post-graduation. This post-survey was purposely placed at the very end to ensure that these questions would not prime the control group as well. For more details about the experiment design, refer back to Section IV: Experiment Trial #. For the full version of this survey, see Appendix C.

B. Data

151 undergraduate students participated in this experiment. However, due to inconsistent and/or missing answers and participants dropping out of the survey, whether it is because of time constraints or lack of incentive, the data was cleaned to just 118 participants. Because there is a smaller sample size this time around, there were also fewer first-gen students (25, compared to 93 non-first-gen). Regardless, these data points are still valuable in adding to our original sample and checking the robustness of the results and whether or not the statistical significance that we originally saw still holds. Jumping ahead, some of the results that were statistically significant in the first trial ended up not being statistically significant in the second trial, which goes to show how complicated it is to isolate the effect of parental education, which requires finding first-gen students and non-first gen students who are exactly the same except for solely their generation status. However, this second trial and the overall experiment design are still meaningful because they suggest that a lab experiment can be done to better isolate the differences between first-gen and non-first-gen and that non-cognitive behaviors, not just socioeconomic factors that we cannot control, may contribute to their economic outcomes. As before, approximately half of these participants were randomly assigned into the treatment group and the other half into the

control group, which resulted in 14 treated first-gen, 11 control first-gen, 45 treated non-first-gen, and 48 control non-first-gen.

C. Patience Game

The new patience game that all participants had to take, regardless of whether they were in the treatment or control group, is the following:

A person randomly stops you on the streets and offers to give you either $5 today or $x in one month. Let's say that this person will definitely follow through on their promise. Which of the following options would you choose?

	$5 today	$x in one month
$5 today or $10 in one month	○	○
$5 today or $15 in one month	○	○
$5 today or $20 in one month	○	○
$5 today or $25 in one month	○	○
$5 today or $30 in one month	○	○

Figure 16: Patience Game. The participants are asked whether they want to receive $5 today or $5+*x* in 7 months. This *x* increases as they go down the list.

The difference between this new patience game and the old one is that the magnitude of the prices offered are much less, ranging from $5 to $30 instead of $10 to $50. This new range allows us to see how the average first-generation college student would react, compared to the average non-first-generation, if they would receive a much lower amount much sooner. After all, the $50 from the old game is a lot, and most college students would love that.

As before, this game was created to measure the participants' patience levels. If the participant is extremely impatient, they would take the $5 today, no matter what the value being offered in a month would be. If they were extremely patient, they would take the $5+*x* in a month, no matter the value of *x*. The point at which a participant switches from choosing $5 today to $\$x$ in one month helps measure their patience level. The sooner the participant makes the switch, the more patient they are.

One way this game was analyzed was by running a Two-Sample Test for Equality. By subsetting the gen groups, we find that first-gens have a mean patience score of 5.68 while non-first-gens have a mean of 5.48. This tells us that first-gens seem to be more patient, which supports the findings from Trial #1 and contradicts my hypothesis. The p-value for this is 0.03019, which shows that this is indeed a significance at the 10% and 5% confidence levels but not at 1%.

When the groups are subsetted by treatment versus control, the mean patience score for the control group is 5.57 while the mean for the primed is 5.43. The p-value for this test is 0.04398, which shows that priming the average participant to make them think about first-gen students, regardless of their gen status; makes them less patient on average, supporting my hypothesis. This is significant at the 10% and 5% confidence levels but not at 1%.

Another way this game was analyzed was by running a regression. When adding both first-gen status and treatment type into the following regression model, however, the results become less strong:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = patience$ (a score determined by the Patience Game),

$x_1 = first\ gen$ (1 if first-gen, 0 if non-first-gen), and

$x_2 = primed$ (1 if primed, 0 if controlled).

Table 1: Patience Game 1

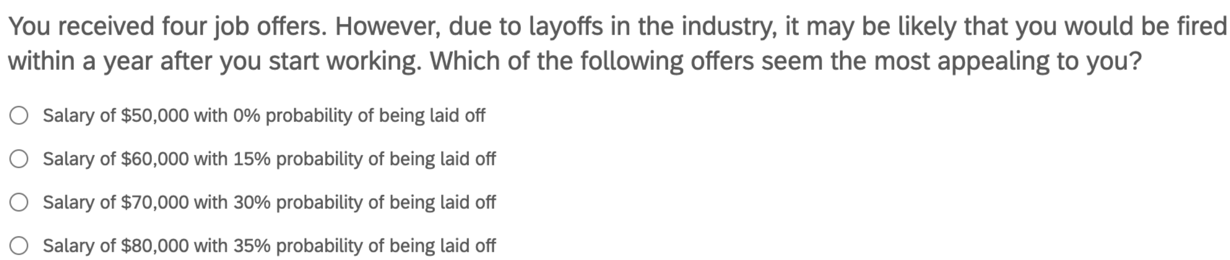
	Dependent variable:
	Patience
First-gen	−0.080
	(0.317)
Primed	−0.307
	(0.198)
First gen and primed	0.547
	(0.431)
Constant	5.625***
	(0.137)
Observations	117
R^2	0.032
Adjusted R^2	0.006
Residual Std. Error	0.950 (df = 113)
F Statistic	1.226 (df = 3; 113)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 17: Regress *Patience* on *First Gen* and *Primed.* These results are unfortunately not statistically significant.

Unfortunately, unlike the Two-Sample Test, these regression results are not statistically significant, most likely due to low power and sample size. But if they were, these results would have supported my hypothesis and the earlier findings because they suggest that first-gens are less patient than non-first-gens and that priming them makes them even more patient. This experiment should be conducted again, but in a controlled lab setting and with more people, to potentially have more accurate and substantial results.

D. Risk Game #1

To elicit their risk preferences, participants were placed in a scenario in which they had to choose one of four job offers, all with varying salaries and lay off rates. If the participant were extremely risk averse, they would choose the first offer since the probability of being laid off is zero, thus guaranteeing the salary. If the participant were extremely risky, they would choose the fourth offer since the probability of being paid off is high, but the participant is willing to take this risk to achieve the higher salary. The risk here is not getting paid, spending months being unemployed, and using a lot of time and energy to find another job. This version, shown below, is simpler and has different numbers than the first trial, but it follows the same concept:


You received four job offers. However, due to layoffs in the industry, it may be likely that you would be fired within a year after you start working. Which of the following offers seem the most appealing to you?

- Salary of $50,000 with 0% probability of being laid off
- Salary of $60,000 with 15% probability of being laid off
- Salary of $70,000 with 30% probability of being laid off
- Salary of $80,000 with 35% probability of being laid off

Figure 18: Risk Game #1. The participant is placed into a scenario where they have to choose one of four job offers with different salaries but also lay off rates.

One way in which this game was analyzed was by running a Two-Sample Test for Equality. After subsetting the generation groups, first-gens have a mean risk score of 2.88 while non-first-gens have a mean of 2.83. This, however, is not statistically significant, due to its p-value of 0.2192.

This insignificance holds when both *first-gen* and *primed* are added into the following regression model:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = risk1$ (a score determined by Risk Game #1),

$x_1 = first\ gen$ (1 if first-gen, 0 if non-first-gen), and

$x_2 = primed$ (1 if primed, 0 if controlled).

Table 2: Risk Game 1

	Dependent variable:
	Risk1
First-gen	−0.134
	(0.349)
Primed	0.118
	(0.217)
First-gen and Primed	0.317
	(0.473)
Constant	2.771***
	(0.151)
Observations	118
R^2	0.012
Adjusted R^2	−0.014
Residual Std. Error	1.045 (df = 114)
F Statistic	0.471 (df = 3; 114)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 19: Regress *Risk1* on *First-gen* and *Primed*. These results are unfortunately not statistically significant.

Unfortunately, as before, these regression results are not significant, most likely due to low power and sample size. But if they were, these results would've supported my hypothesis because they suggest that first-gens may be less risky than non-first-gens, holding all else constant.

E. Risk Game #2

Another risk game is the following:

You are back on the job market again. You currently already have a job offer that pays $50,000 that you need to respond to within three weeks. You could accept the offer and not have to worry about recruiting for the rest of the year, or you could decline the offer and re-recruit, with no guarantee that you would find something better. If you turn down the offer, there is some chance that you will not have a job when you graduate from college and some chance that you will have a job offer that pays $100,000. In each of the rows below, the probability of getting the higher-paying job varies.

	Accept Job Offer	Re-recruit
A	$50,000 guaranteed	10% chance of $100,000
B	$50,000 guaranteed	20% chance of $100,000
C	$50,000 guaranteed	30% chance of $100,000
D	$50,000 guaranteed	40% chance of $100,000
E	$50,000 guaranteed	50% chance of $100,000
F	$50,000 guaranteed	60% chance of $100,000

Which option would you choose?

- ○ A
- ○ B
- ○ C
- ○ D
- ○ E
- ○ F

Figure 20: Risk Game #2. The participant is placed into a scenario in which they have to choose between accepting a job offer that pays a lower salary or declining it to re-recruit for the chance to obtain a higher salary.

This situation is very similar to the previous game, except this time, the participant is re-recruiting with no guarantee of landing a job. This is a common real-life situation that many college students have to go through, but I wonder if first-generation college students make different choices from their non-first-generation peers.

One way this game was analyzed was by running a Two-Sample Test for Equality. After subsetting the gen groups, we find that first-gen participants have a mean risk score of 1.38 while that for non-first-gen is 1.8. Unfortunately, this difference is not statistically significant, due to the p-value being 0.3836, but it would have supported my hypothesis.

This insignificance holds when we add both *first-gen* and *primed* into the following regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \text{, where}$$

$y = risk2$ (a score determined by Risk Game #2),

$x_1 = first\ gen$ (1 if first-gen, 0 if non-first-gen), and

$x_2 = primed$ (1 if primed, 0 if controlled).

Table 3: Risk Game 2

	Dependent variable:
	Risk2
First-gen	0.080
	(0.383)
Primed	0.003
	(0.238)
First-gen and primed	0.614
	(0.519)
Constant	1.375***
	(0.165)
Observations	118
R^2	0.038
Adjusted R^2	0.013
Residual Std. Error	1.145 (df = 114)
F Statistic	1.495 (df = 3; 114)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 21: Regress *Risk* on *First-gen* and *Primed.* These results unfortunately are not statistically significant.

The regression results suggest that first-gens may be more risky than non-first-gen, holding all else constant, and that primed first-gens are even riskier, which both contradict our hypothesis and existing literature. This is not a problem because the results are actually insignificant, due to the high p-values, but it is still important to note that the definition of risk may vary from person to person and that it is a hard thing to quantify. Some argue that first-gens are less risky than non-first-gens (Trejo, 2016) because they elect into majors that have a set path for them. However, one might say that first-gens are actually more risky: They are the trailblazers in their family, walking on a path that was not set out for them by their parents. As you can see, the definition of risk aversion varies from person to person, and there is no one way to measure it, which may have led to some of these contradictory results.

F. Aggregate Risk Score

The risk scores from both of the previous games were aggregated to gain a better understanding of the overall risk levels of first-generation and non-first-generation college students. The individual effects themselves may be statistically insignificant, but adding these small effects together may actually make the overall effect be significant, which does end up happening.

The aggregated results tell us that the average aggregated risk score for first-gens was 4.68 while that for non-first-gens was 4.20. A Two-Sample Test for Equality produces a p-value of 0.04964, which can be rejected at the 5% significance level. This tells us that there is a 95% chance that first-gens are statistically riskier than non-first-gens, which again, contradicts some literature, but it depends on how one defines risk. For example, one can make the argument that first-gens are less risky because they tend to choose majors with defined career paths and little to no uncertainty about finding a job post-graduation (Trejo, 2016). But one may argue that by being trailblazers and walking on a path that no one else in their family has done before, first-gens are making a risky decision in the first place. Regardless of what the true measure of risk may be, this experiment shows that there is indeed a difference here. The definition of this exact difference is open to interpretation and has room to be further explored in future experiments.

As before, however, the effects are less strong once both the *first-gen* and *primed* are placed together, but they still give the following interesting results:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$y = aggregate\ risk$ (the mean across Risk Games 1-2),

$x_1 = first\ gen$ (1 if first-gen, 0 if non-first-gen), and

$x_2 = primed$ (1 if primed, 0 if controlled).

Table 4: Aggregate Risk Score

	Dependent variable:
	Aggregate Risk
First-gen	−0.055
	(0.527)
Primed	0.121
	(0.327)
First-gen and Primed	0.931
	(0.715)
Constant	4.146***
	(0.228)
Observations	118
R^2	0.039
Adjusted R^2	0.014
Residual Std. Error	1.577 (df = 114)
F Statistic	1.557 (df = 3; 114)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 22: Regress *Aggregate Risk* on *First-gen* and *Primed*. Unfortunately, these results are not statistically significant.

Unfortunately, these regression results are not statistically significant, but it would've supported my hypothesis that first-gens are less risky. Surprisingly, though, priming first-gens makes them riskier, perhaps explaining the high risk scores from earlier. This, again, however, may just be symbolic of the fact that different people may define risk differently.

G. Confidence

As an addition to the first, this trial looked into whether or not there is a difference in the confidence levels between first-generation and non-first-generation college students. The reason behind exploring confidence is that it may be a contributor to the differences in economic outcomes that we see between first-gens and non-first-gens, as shown in the study by the Pew

Research Center (2021) and in my SHED Analysis from Section III. Is it that certain employers and industries discriminate against first-gens and disallow them from entering certain companies and industries? This is unlikely because the majority of applications don't ask about first-gen status, let alone family history. Or is it that first-gens are not electing themselves into these employers and industries, in the first place, due to their potential lack of confidence?

This question is a large one, and there could be an entire experiment devoted into whether or not first-gens are less confident than non-first-gens and whether or not first-gens would select different careers, roles, job offers, companies; etc. than non-first-gen students when they are told to apply to hypothetical job postings given hypothetical resumes.

But in an attempt to answer this, this experiment revised a question that is commonly asked in surveys:

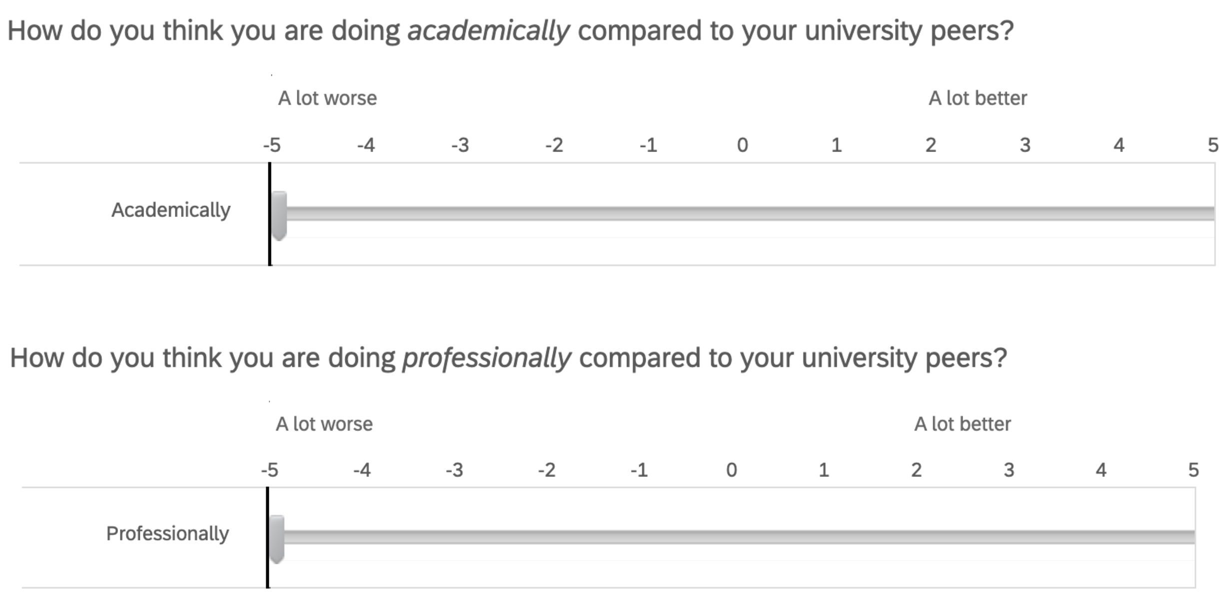


Figure 23: These questions gauged the participants' confidence levels by asking them to compare themselves to their peers.

This common question, most famously used by Boneva et al (2021), is an interesting one because it gauges confidence levels, without explicitly asking the participants to rate their self-confidence levels themselves. Doing the latter could result in biases that could skew our results.

From the first question, the mean academic confidence score for first-gens was 1.08 while that for non-first-gens was 1.34. From the second question, the mean professional confidence score for first-gens was 0.52 while that for non-first-gens was 0.677. Unfortunately, these differences are statistically insignificant, and these insignificances persist in their regression forms as well and when aggregated. In other words, there doesn't seem to be major takeaways from these confidence questions, but research in this area should not be discouraged. Only two questions were asked in this category, which may not be enough to truly measure confidence levels, and perhaps there are better questions and better games to implement. Another round of experiments that focuses on solely this topic should be done.

H. Competitiveness

Last but not least, another addition to the first is that this trial looked into whether or not competitiveness is another difference -- that is both cognitive and non-cognitive -- between first-generation and non-first-generation college students, which may contribute to their differences in key economic outcomes (Pew Research Center, 2021). Similar to before, maybe the reason most non-first-gen college graduates are in "private-for-profit" companies in the "Professional, Scientific, Technical, and Business Services" industries (SHED, 2019) is because they are more competitive and thus, are more willing to be in these competitive pools of applicants.

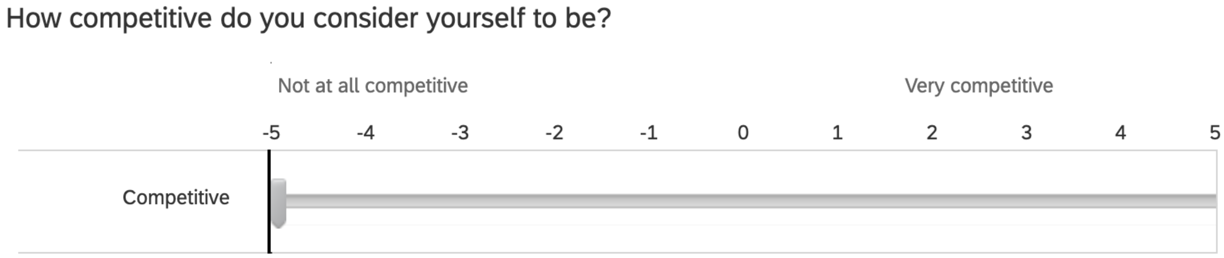


Figure 24: This question asks the participant to rate their competition level themselves.

From a Two-Sample Test for Equality, there does indeed seem to be a significant difference in the competitiveness levels between first-gens and non-first-gens, but perhaps not in the way that one would expect: The average self-recorded competitiveness level for first-gens in our sample is 6.08 while that for non-first-gens is 5.78. The p-value from this is 0.02155, which can be rejected at the 5% significance level. This means that there is a 95% chance that first-gens have a mean competitive score of 6.08, which is more than the 5.78 for non-first-gens.

It is important to note here, however, that the first-gen students in this sample is not representative of all first-gen students as a whole, and thus, one cannot generalize this finding to all first-gen students. The first-gen students here, from renowned American universities such as the University of Michigan, University of Wisconsin, and Boston University; might be more competitive than most people in general because these are competitive schools to get into in the first place, with University of Michigan -- Ann Arbor's acceptance rate being 20.2% (IPEDS, 2021). These first-gen students, if they are aware of their barriers to enter college, may even feel the need to do even more than others in order to be accepted, making them even more competitive.

And as before, this game consisted of only one question that is self-reported, and so, it may not be the best measure of confidence. Thus, an entire experiment on its own can -- and should -- be run in this area, such as the infamous one by Niederle and Vesterlund (2007), which examines whether men or women are more likely to elect into a competitive compensation (only get paid a large amount if they answered the most questions correctly out of the group) or a piece-rate compensation (get paid a steady amount per question answered correctly).

Conclusion

In conclusion, through an analytical and experimental approach, this paper explored potential reasons for the differences in the key economic outcomes between first-generation and non-first-generation college students. The analytical approach using the SHED (2020) data found that non-first-gen college graduates tend to work for "private-for-profit" companies in the "Professional, Scientific, Technical, and Business Services" industry, which tend to pay more, so this may contribute to the outcome differences.

Meanwhile, the experimental approach found that first-gen students were less patient than their counterparts, which supports my hypothesis and definition of patience that they are less likely to stay for school for longer (to pursue a Doctorate or even Master's degree).

The experimental approach, however, had conflicting results regarding riskiness. Some risk games suggest that first-gen students are less risky, which supports past literature (Trejo, 2016) and my hypothesis, but others suggest that first-gen students have a higher aggregate risk score, which one may argue that may be the case because first-gens are making a risky decision in the first place by being trailblazers and walking on a path that no one else in their family has done before.

Unfortunately, the experimental approach did not seem to produce significant differences between first-gens and non-first-gens in terms of confidence, but there may be some significances there if more, and even better, questions were asked.

Last but not least, the experimental approach suggests that there is a significant difference in the competitiveness levels between first-gen and non-first-gen students, but perhaps not in the way that one would suspect: First-gen students have a higher self-recorded competitiveness level. As mentioned before, however, this cannot be generalized to the entire first-gen student population, as the first-gen students in the sample may be more competitive than average, considering the fact that they attend universities that are competitive to get into in the first place.

In the future, instead of spreading so thin, research and experiments can -- and should -- be done on any one of these non-cognitive behaviors (patience, risk aversion, confidence, and competitiveness) to further explore potential and significant differences. An experiment can be done, for example, to see whether first-gen and non-first-gen students are different in terms of competitiveness, specifically. Studies like this paper, and hopefully future research, can help find the differences between first-generation and non-first-generation college students early on, and this is relevant because it is only when we know these differences that one can attempt to bridge the gaps that we see in economic outcomes later on.

References

Algan, Yann, Pierre Cahuc, Andrei Shleifer, and Markus Wilhelm. 2022. “The Impact of Childhood Social Skills and Self-Control Training on Economic and Noneconomic Outcomes.” *Journal of Political Economy*, forthcoming.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." *American Economic Review*, 100 (4): 1913-28.

Boneva, Teodora, Thomas Buser, Armin Falk, and Fabian Kosse. 2021. "The Origins of Gender Differences in Competitive and Earnings Expectations: Causal Evidence from a Mentoring Intervention." *Journal of Political Economy*, 129 (4): 1064-1104.

Bureau of Labor Statistics. n.d. "Occupational Employment and Wages, May 2021." Accessed December 13, 2022. https://www.bls.gov/oes/current/oes_nat.htm.

Choy, Susan P. 2001. Students Whose Parents Did Not Go To College: Postsecondary Access, Persistence, and Attainment (NCES 2001-126). *Washington, DC: U.S. Department of Education, National Center for Education Statistics*. http://nces.ed.gov/pubs2001/2001126.pdf.

Emmons, William R., Ana Hernandez Kent, and Lowell Ricketts. 2021. "First-Generation College Graduates Get a Financial Boost, but Don’t Catch Up." *Federal Reserve Bank of St. Louis*, December 9. Accessed April 25, 2022. https://www.stlouisfed.org/publications/in-the-balance/2019/first-generation-college-graduates.

Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. "Estimating Individual Discount Rates in Denmark: A Field Experiment." *American Economic Review* 92 (5): 1606-1617.

Kent, Ana Hernandez. 2021. "Children of College Graduates Earn More and Are Richer." *Federal Reserve Bank of St. Louis.* December 9. Accessed April 25, 2022. https://www.stlouisfed.org/publications/in-the-balance/2019/children-of-college-graduates.

Leukhina, Oksana, and Andrew Smaldone. 2022. "Why Do Women Outnumber Men in College Enrollment?" *Federal Reserve Bank of St. Louis.* March 14. Accessed December 13, 2022. https://www.stlouisfed.org/on-the-economy/2022/mar/why-women-outnumber-men-college-enrollment.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics 122* (3): 1067-1101. https://doi.org/10.1162/qjec.122.3.1067.

Pew Research Center. 2021. "First-Generation College Graduates Lag Behind Their Peers on Key Economic Outcomes." May.

The Federal Reserve. 2020. "Survey of Household Economics and Decisionmaking." *The Federal Reserve.* Accessed April 25, 2021. https://www.federalreserve.gov/consumerscommunities/shed.htm.

Trejo, Sam. 2016. An Econometric Analysis of the Major Choice of First-Generation College Students. *The Developing Economist 3* (1), http://www.inquiriesjournal.com/a?id=1407

William, James. 1890. *The Principles of Psychology*. New York: Henry Holt and Company. doi: http://dx.doi.org/10.1037/11059-000.

Appendix A: Experiment Trial #1 Qualtrics Survey

Default Question Block

This is a study to help us better understand the undergraduate population at the University of Michigan. Your responses will be tracked. There are no right or wrong answers. Please answer the following questions to the best of your abilities. This survey will take approximately 10 minutes to complete. You will receive a $2-3 gift card for your participation.

Demographic

First, please answer some demographic questions.

What is your email?

What is your gender?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

What is your race?

- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or Pacific Islander
- ☐ Other

What is your year in school?

- ○ 1st year
- ○ 2nd year
- ○ 3rd year
- ○ 4th year
- ○ 5th year

What is[are] your major[s]?

____

Priming

What is your mother's highest education level?

- ☐ Some high school
- ☐ High School Diploma/GED
- ☐ Some College
- ☐ Bachelors
- ☐ Masters
- ☐ PhD/MD
- ☐ Associate's
- ☐ Other

What is your father's highest education level?

- ☐ Some high school
- ☐ High School Diploma/GED
- ☐ Some College
- ☐ Bachelors
- ☐ Masters
- ☐ PhD/MD
- ☐ Associate's
- ☐ Other

What proportion of students at the University of Michigan do you believe are first-generation college students, meaning that they are the first in their immediate family to attend college? 0-100%

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Proportion of students that are first-generation											

How much support did you get from your family in applying for/going to college?

- ○ None
- ○ Little Support
- ○ Moderate Support
- ○ A lot of Support

Do you think college is more or less challenging for first generation college students?

- ○ Much less challenging
- ○ Less challenging
- ○ Neither more or less challenging
- ○ More challenging
- ○ Much more challenging

Control

Which of the following streaming services do you use?

- ☐ Netflix
- ☐ Hulu
- ☐ Peacock
- ☐ HBO Max
- ☐ Disney Plus
- ☐ Amazon Prime
- ☐ Paramount

Which campus library do you utilize the most?

- ○ Hatcher
- ○ Shapiro
- ○ Law
- ○ Duderstadt

How many course credits are you taking this semester?

____

Block 7

In this next section, you will be presented a series of hypothetical scenarios.

Time

One of your friends is kind and is willing to give you a one-time payment. However, you may only receive the payment either 1 month later or 7 months later. You will be compensated more if you choose the delayed payment option.

For each row, choose between option A or B.

	Payment Option	
	A (1 month)	B (7 months)
$10 in 1 month or $15 in 7 months	○	○
$10 in 1 month or $20 in 7 months	○	○

	Payment Option	
	A (1 month)	B (7 months)
$10 in 1 month or $25 in 7 months	○	○
$10 in 1 month or $30 in 7 months	○	○
$10 in 1 month or $35 in 7 months	○	○
$10 in 1 month or $40 in 7 months	○	○
$10 in 1 month or $45 in 7 months	○	○
$10 in 1 month or $50 in 7 months	○	○

Risk Preferences

Congrats, you've made it to the final round in a popular game show! You may go home with $10,000 guaranteed if you choose, but if you participate in the challenge round, you may have the opportunity to win even more: $20,000! For each of the following chances, would you accept or decline the challenge round?

	Accept/Decline Challenge R	
	Accept	Decl
Accept 10% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 20% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 30% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 40% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 50% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 60% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 70% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 80% chance of $20,000 or Decline with $10,000 guaranteed	○	○
Accept 90% chance of $20,000 or Decline with $10,000 guaranteed	○	○

You are a sales associate for a small boutique, and you need to decide which days of the week that you have to go in for work. Because your job is commissions based, however, there is no guarantee that you will get the full amount. The following chart shows the potential payoffs from each shift schedule. Which of the following shift schedules would you choose?

- ½ of $80, ½ of $12
- ½ of $70, ½ of $22
- ½ of $62, ½ of $26
- ½ of $54, ½ of $30
- ½ of $46, ½ of $34
- ½ of $38, ½ of $38

You received multiple job offers. As you go down your list of job salaries, the salaries significantly increase, from as little as $40,000 to as much as $70,000. However, as the salary increases, the probability that you get laid off within the first few months (and thus, do not get paid the full salary) increases as well. You can only accept one job offer at this time. Which of the following job offers would you choose?

- Salary of $40,000 with 0% probability of being laid-off
- Salary of $75,000 with 10% probability of being laid-off
- Salary of $100,000 with 20% probability of being laid-off
- Salary of $200,000 with 30% probability of being laid-off

Post

Are you a first-generation college student?

- Yes
- No

How confident are you that you will complete your degree?

- 1 (Not confident)
- 2
- 3

- 4
- 5 (Very confident)

How many total years do you expect to need to complete your degree?

- 2
- 2.5
- 3
- 3.5
- 4
- 4.5
- 5
- 5.5
- 6

What are your *expected* annual earnings in your first job post-graduation? (Please round to the nearest thousand)

What are your *desired annual* earnings in your first job post-graduation? (Please round to the nearest thousand)

What are your expected plans post-graduation?

- Work full-time
- Graduate School
- Gap Year
- Research
- Other

Indicate on a scale from 1 (completely disagree) to 5 (completely agree) on how likely you are to agree/disagree on the following statements.

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	
	0	1	3	4	5	
Going to college is a risky investment						
Going to college will guarantee you a lot of money in the future						
You don't need to go to college to be successful						
Going to college is a worthy investment even if it is expensive and may require me to take on loans						
Going to college will help improve job stability						
I need to go to college to achieve my future goals						
Taking on student loans is too risky to justify going to college						

Appendix B: Experiment Trial #1 Code

```
#------------------------
# Econ 408 Final Project
# Version 01_sharon
#------------------------

#--------------------
# 1a: Set Directory
#--------------------
#setwd("~/Downloads")
library(tidyverse)
library(tidyr)
library(haven) #to import data
library(dplyr) #to filter
library(ggplot2)
library(stargazer)
library(lemon)
library(knitr)
library(kableExtra)
library("ggpubr")

#------------------
# 1b: Import Data & Clean
#------------------
survey<-read.csv("survey.csv",encoding="UTF-8")

#Remove extraneous columns and rows
survey <-  survey[ -c(1:17) ]
survey<- survey[-2,]

survey_w_question <- survey
```

```r
survey <-survey[-1,]

#factorize variables
survey[, c(2:4,6,7,9,10,14:35,38)] <- lapply(survey[, c(2:4,6,7,9,10,14:35,38)], as.factor)

#convert continuous variables to numeric
survey[, c(8,36:37,39:45)] <- lapply(survey[, c(8,36:37,39:45)], as.numeric)

#remove responses where they did not disclose their first gen status
n_removed <- survey %>% filter(Q20.1=="") %>% nrow()
survey <- survey %>% filter(!Q20.1=="", !Q3=="")

table(survey$student_status)
table(survey$student_group)

#-------------------
# Part 1c: Subsets
#-------------------

prop.table(table(survey$Q7,survey$Q20.1))
table(survey$Q7, survey$Q8)


# Make a variable name for treatment/control

survey <- survey %>% mutate(group=ifelse(Q16=="A lot of Support"|
                                  Q16=="Little Support"|
                                  Q16=="Moderate Support"|
                                  Q16=="None","Treatment", "C
survey$student_group <- as.factor(paste(survey$group, survey$student_status))

# Create the treatment group (the group of first-gen students who were primed):
# Those who said yes to q20.1
# AND those who answered q16
primed_first_gen<-filter(survey,Q20.1=="Yes" & (Q16=="A lot of Support"|
                                  Q16=="Little Support"|
                                  Q16=="Moderate Support"|
                                  Q16=="None"))

# Create the control group (the group of first-gen students who were NOT primed):
# Those who said yes to q20.1
# AND those who answered q18
control_first_gen<-filter(survey,Q20.1=="Yes" & (Q19=="Duderstadt"|Q19=="Hatcher"|
                                  Q19=="Law"|Q19=="Shapiro"))

# Check: total obs of groups above approx add up to total obs for following:
first_gen<-filter(survey,Q20.1=="Yes")

# Create the primed non-first-gen group:
# Those who said no to q20.1
```

```
# AND those who answered q16
primed_non_first_gen<-filter(survey,Q20.1=="No" & (Q16=="A lot of Support"|
                                                  Q16=="Little Support"|
                                                  Q16=="Moderate Support"|
                                                  Q16=="None"))

# Create the non-primed non-first-gen group:
# Those who said no to q20.1
# AND those who answered q18
control_non_first_gen<-filter(survey,Q20.1=="No" & (Q19=="Duderstadt"|Q19=="Hatcher"|
                                                   Q19=="Law"|Q19=="Shapiro"))

# Check: total obs of groups above approx add up to total obs for following:
non_first_gen<-filter(survey,Q20.1=="No")

#--------------------
# 1d: Summary Stats
#--------------------

summary(survey$Q7) #134 female, 83 male, 4 other
summary(first_gen$Q7) #26 female, 17 male, 1 other
summary(non_first_gen$Q7) #108 female, 66 male, 3 other

summary(survey$Q8) #96 Asian, 3 Black, 7 Other, 107 White, 8 White/Asian
summary(first_gen$Q8) #19 Asian, 1 Black, 4 Other, 20 White
summary(non_first_gen$Q8) #77 Asian, 2 Black, 3 Other, 87 White, 8 White/Asian

summary(survey$Q9) #17 fresh, 34 soph, 77 jr, 90 snr, 1 fifth yr
summary(first_gen$Q9) #7 fresh, 8 soph, 10 jr, 18 snr, 1 fifth yr
summary(non_first_gen$Q9) #10 fresh, 26 soph, 67 jr, 72 snr, 2 fifth yr

#------------------
# Patience Game
#------------------

# what is the proportion of primed first gen kids who choose a over b (impatient)

prop.table(table(primed_first_gen$Q21.1_1)) # 74% of primed first-gen are impatient, 26% patie
prop.table(table(control_first_gen$Q21.1_1)) # 75% of control first-gen are impatient, 25% pat
                                            # no sig diff b/w primed and control first-gen

prop.table(table(primed_first_gen$Q21.1_2)) # 61% impatient, 39% patient
prop.table(table(control_first_gen$Q21.1_2)) # 60% impatient, 40% patient
                                            # no sig diff

prop.table(table(primed_first_gen$Q21.1_3)) # 48% impatient, 52% patient
prop.table(table(control_first_gen$Q21.1_3)) # 40% impatient, 60% patient
                                            # sig diff

prop.table(table(primed_first_gen$Q21.1_4)) # 26% impatient, 74% patient
```

```
prop.table(table(control_first_gen$Q21.1_4)) # 30% impatient, 70% patient
                                            # sig diff

prop.table(table(primed_first_gen$Q21.1_5)) # 13% impatient, 87% patient
prop.table(table(control_first_gen$Q21.1_5)) # 10% impatient, 90% patient
                                            # sig diff

prop.table(table(primed_first_gen$Q21.1_6)) # 9% impatient, 91% patient
prop.table(table(control_first_gen$Q21.1_6)) # 10% impatient, 90% patient
                                            # no sig diff

prop.table(table(primed_first_gen$Q21.1_7)) # 4% impatient, 96% patient
prop.table(table(control_first_gen$Q21.1_7)) #10% impatient, 90% patient
                                            # sig diff

primed_first_gen<-filter(primed_first_gen,Q21.1_8=="A (1 month)"|Q21.1_8=="B (7 months)")
prop.table(table(primed_first_gen$Q21.1_8)) # 4% impatient, 91% patient
prop.table(table(control_first_gen$Q21.1_8)) # 5% impatient, 95% patient



prop.table(table(primed_non_first_gen$Q21.1_1)) # 53% of primed non-first-gen are impatient,
prop.table(table(control_non_first_gen$Q21.1_1)) # 68% of control non-first-gen are impatient,
                                                # sig diff

primed_non_first_gen<-filter(primed_non_first_gen,Q21.1_2=="A (1 month)"|Q21.1_2=="B (7 months
control_non_first_gen<-filter(control_non_first_gen,Q21.1_2=="A (1 month)"|Q21.1_2=="B (7 mont
prop.table(table(primed_non_first_gen$Q21.1_2)) # 42% impatient, 52% patient
prop.table(table(control_non_first_gen$Q21.1_2)) # 45% impatient, 53% patient
                                                # sig diff

primed_non_first_gen<-filter(primed_non_first_gen,Q21.1_3=="A (1 month)"|Q21.1_3=="B (7 months
control_non_first_gen<-filter(control_non_first_gen,Q21.1_3=="A (1 month)"|Q21.1_3=="B (7 mont
prop.table(table(primed_non_first_gen$Q21.1_3)) # 26% impatient, 67% patient
prop.table(table(control_non_first_gen$Q21.1_3)) # 31% impatient, 68% patient
                                                # sig diff

prop.table(table(primed_non_first_gen$Q21.1_4)) # 17% impatient, 77% patient
prop.table(table(control_non_first_gen$Q21.1_4)) # 20% impatient, 80% patient
                                                # sig diff

prop.table(table(primed_non_first_gen$Q21.1_5)) # 7% impatient, 88% patient
prop.table(table(control_non_first_gen$Q21.1_5)) # 11% impatient, 86% patient
                                                # sig diff

prop.table(table(primed_non_first_gen$Q21.1_6)) # 6% impatient, 89% patient
prop.table(table(control_non_first_gen$Q21.1_6)) # 5% impatient, 93% patient
                                                # no sig diff

prop.table(table(primed_non_first_gen$Q21.1_7))  # 5% impatient, 90% patient
```

```
prop.table(table(control_non_first_gen$Q21.1_7)) # 5% impatient, 93% patient
                                                # no sig diff
primed_non_first_gen<-filter(primed_non_first_gen,Q21.1_8=="A (1 month)"|Q21.1_8=="B (7 months
prop.table(table(primed_non_first_gen$Q21.1_8)) # 3% impatient, 97% patient
prop.table(table(control_non_first_gen$Q21.1_8)) # 2% impatient, 95% patient
                                                # no sig diff

#--------------------
# Risk Pref Game #1
#--------------------

risk_1 <- survey[ 22:30]


risk_1 <- risk_1 %>%
  mutate(cutoff=ifelse(Q28.1_1!=Q28.1_2, 1,
                       ifelse(Q28.1_2!=Q28.1_3,2,
                              ifelse(Q28.1_3!=Q28.1_4,3,
                                     ifelse(Q28.1_4!=Q28.1_5,4,
                                            ifelse(Q28.1_5!=Q28.1_6,5,
                                                   ifelse(Q28.1_6!=Q28.1_7,6,
                                                          ifelse(Q28.1_7!=Q28.1_8,7,
                                                                 ifelse(Q28.1_8!=Q28.1_9,8,9))

survey$risk_game_1_cutoff <-risk_1$cutoff
survey$risk_game_1_cutoff <-10-survey$risk_game_1_cutoff

#higher score means more risky

survey %>% group_by(student_status) %>% summarize(mean_risk_game_1_cutoff=mean(risk_game_1_cu
t.test(risk_game_1_cutoff~student_status, data=survey)
reg1<-lm(risk_game_1_cutoff~group+student_status+(group*student_status),data=survey)

#----------------------------
# Part 3b: Risk Pref Game #2
#----------------------------
# Create variable for risk:

# Convert character vector to factor
survey$Q29_numeric <- as.numeric(factor(survey$Q29, levels = as.character(unique(survey$Q29)))
reg2<-lm(Q29_numeric~group+student_status+(group*student_status),data=survey)

summary(reg2)

#higher score means more risky

t.test(Q29_numeric~student_status, data=survey)

#--------------------
# Risk Pref Game #3
```

```
#--------------------

# Create risk variable (1 being least, 4 being highest):
survey$riskgame3<-0
survey$riskgame3[survey$Q32=="Salary of $40,000 with 0% probability of being laid-off"]<-1
survey$riskgame3[survey$Q32=="Salary of $75,000 with 10% probability of being laid-off"]<-2
survey$riskgame3[survey$Q32=="Salary of $100,000 with 20% probability of being laid-off"]<-3
survey$riskgame3[survey$Q32=="Salary of $200,000 with 30% probability of being laid-off"]<-4

# Checks:
table(survey$Q32)
table(survey$riskgame3)

# Regress group, student status, and interaction term on risk:
reg3<-lm(riskgame3~group+student_status+(group*student_status),data=survey)
summary(reg)

#----------------------------
# Part 3d: Add up risk preference scores
#----------------------------

survey$risk_score <- (survey$riskgame3+survey$risk_game_1_cutoff+survey$Q29_numeric)

#a higher score correlates to being more risky

#----------------------------
# Part 4: Regression Analysis
#----------------------------

risk_mod <- lm(risk_score~group+student_status+group*student_status, data=survey)
summary(risk_mod)
t.test(risk_score~student_status, data=survey) #first gen students statistically significantly

boxplot(risk_score~student_status*group, data=survey, col="light blue")

stargazer(reg1, reg2, reg3,risk_mod, dep.var.labels= c("Risk Game 1", "Risk Game 2", "Risk Gam

#There is a close statistically significant effect of priming in sentiment score for first-ger
t.test(survey$risk_score[survey$student_status=="First-gen"]~survey$group[survey$student_statu

#There is no statistically significant effect of priming in sentiment score for non first-gen
t.test(survey$risk_score[survey$student_status=="Not First-gen"]~survey$group[survey$student_s

risk_table <- survey %>% group_by(student_group) %>% summarize( mean_risk1=mean(risk_game_1_cu

risk_table %>%
  kbl(caption = "Risk Scores by Experimental Group and First-Generation Status", col.names=c('
  kable_classic(full_width = F, html_font = "Cambria")

#----------------------------
```

```
# Part 5: Sentiment Score
#----------------------------

#Reverse scoring and score calculation
reverse_cols <- c("Q26_1", "Q26_3", "Q26_7")
survey[ , reverse_cols] <- 5 - survey[ , reverse_cols]

survey <- survey %>% mutate(sentiment_score = select(., Q26_1:Q26_7) %>% rowSums(na.rm = TRUE)

#difference in the sentiment score for first generation status and group assignment
survey %>% group_by(Q20.1, group) %>% summarize(mean_sentiment_score=mean(sentiment_score))

#There is no statistically significant difference in sentiment score and first generation stat
t.test(survey$sentiment_score~survey$Q20.1)

#There is no statistically significant effect of priming in sentiment score for first-gen stud
t.test(survey$sentiment_score[survey$student_status=="First-gen"]~survey$group[survey$student_

#There is no statistically significant effect of priming in sentiment score for non first-gen
t.test(survey$sentiment_score[survey$student_status=="Not first_gen"]~survey$group[survey$stud

aov1 <- aov(sentiment_score~student_group, data=survey)
summary(aov1)

boxplot(sentiment_score~student_status*group, data=survey, col="light blue")

#Fit a linear regression model

lm1 <- lm(sentiment_score~student_status+group+student_status*group, data=survey)
summary(lm1)
stargazer(lm1, dep.var.labels= c("Sentiment Score"), covariate.labels= c( "Not First-Generatio
out="sentiment.htm")

lm2 <- lm(sentiment_score~risk_score*student_status, data=survey)
summary(lm2)

cor(survey$sentiment_score,survey$risk_score, method = "pearson")

#----------------------------
# Part 6: Exploratory Analysis
#----------------------------

prop.table(table(survey$student_status, survey$Q25))

#within the treatment group, how are risk and time preferences dependent on parents education
library(ggplot2)
survey %>% filter(group=="Treatment") %>% ggplot(aes(x=Q13, y=risk_score, fill=student_status)
survey %>% filter(group=="Treatment") %>% ggplot(aes(x=Q13, y=sentiment_score, fill=student_st

parent_edu_mod <- lm(sentiment_score~Q13*student_status+Q13+student_status, data=survey)
```

```r
summary(parent_edu_mod)

#expected earnings
boxplot(Q23~student_status,data=survey, main="Expected Earnings by First Generation Status",
        xlab="First Generation Status", ylab="Expected Earnings after Graduation")

boxplot(Q23~Q7,data=survey, main="Expected Earnings by Gender",
        xlab="Gender", ylab="Expected Earnings after Graduation")

earnings_mod <- lm(Q23~student_status+group+student_status*group, data=survey)


#do ppl think college is harder for first gen students?
#parental education background
```

Link to the GitHub:

https://github.com/sharonma1218/Econ408FinalProject/blob/9fee38212eb009f0a3b235adcd2c37200cbc7823/01_sharon.R

Appendix C: Experiment Trial #2 Qualtrics Survey

Study Title: College Students Versus Behavior
Principle Investigator: Sharon Ma, Student, LSA Economics
Faculty Advisor: Basit Zafar

You are invited to take part in a research study. This form contains information that will help you decide whether to join this study.

- The purpose of this survey is to better understand the behavior of college students.
- If you choose to participate, you will be asked a series of questions. No prior knowledge is needed to answer any of these questions, and there are no right or wrong answers.
- Any personal identifiers, such as your name or email address, will be deleted at the end of the data collection, and therefore, none of your answers would be traced back to you, protecting your privacy.
- This survey will take no more than 10 minutes to complete, and you do not need anything more than just a device, whether that be a computer or a smartphone.
- You will receive a guaranteed $2-5 gift card for your participation.

Taking part in this research project is voluntary. You do not have to participate, and you can stop at any time.

If you have any questions about this study or about your rights as a research participant, please contact sharoma@umich.edu.

By signing this document, you are agreeing to take part in this study.

Full Name

Email

Street Address (optional if not umich student)

UNIVERSITY OF MICHIGAN

University

- ◯ University of Michigan - Ann Arbor
- ◯ Other

Year in School

- ◯ Freshman
- ◯ Sophomore
- ◯ Junior
- ◯ Senior
- ◯ Master's
- ◯ PhD/MD/JD

Major

Minor (if applicable)

What is your mother's education level?

- Some high school
- High School
- Some college
- Bachelor's
- Master's
- PhD/MD/JD
- Other

What is your father's education level?

- ◯ Some high school
- ◯ High School
- ◯ Some college
- ◯ Bachelor's
- ◯ Master's
- ◯ PhD/MD/JD
- ◯ Other

How much support did you get from your family in applying for/going to college?

- ◯ None at all
- ◯ A little
- ◯ A moderate amount
- ◯ A lot
- ◯ A great deal

How much pressure did you receive from your family to attend college?

- ◯ None at all
- ◯ A little
- ◯ A moderate amount
- ◯ A lot
- ◯ A great deal

Approximately how much is your total household income (your income and/or your parents')? Round to the nearest thousand.

How often does your family worry about paying the bills?

- ◯ Never
- ◯ Sometimes
- ◯ About half the time
- ◯ Most of the time
- ◯ Always

Did you have to take out student loans?

- ◯ Yes
- ◯ No

Approximately how much do you have to pay in tuition out of pocket per semester?

- ◯ <10,000
- ◯ 10,000-20,000
- ◯ 20,000-30,000
- ◯ 30,000-50,000
- ◯ >50,000

Approximately how much of your tuition is covered by need-based financial aid? (0% if no need-based financial aid)

0 10 20 30 40 50 60 70 80 90 100

Proportion of Tuition Covered

How many months after graduation do you believe is an acceptable time to start working full-time?

0 1 2 4 5 6 7 8 10 11 12

Months

If you are a first-generation college student, do you think navigating through college life is more or less challenging, compared to your peers?

- Much less challenging
- Less challenging
- About the same
- More challenging
- Much more challenging
- Not applicable

What proportion of your university's students do you believe are first-generation students (the first in their immediate family to attend college)?

0 10 20 30 40 50 60 70 80 90 100

Proportion of First-Gen

UNIVERSITY OF MICHIGAN

A person randomly stops you on the streets and offers to give you either $5 today or $x in one month. Let's say that this person will definitely follow through on their promise. Which of the following options would you choose?

	$5 today	$x in one month
$5 today or $10 in one month	○	○
$5 today or $15 in one month	○	○
$5 today or $20 in one month	○	○
$5 today or $25 in one month	○	○
$5 today or $30 in one month	○	○

UNIVERSITY OF MICHIGAN

You received four job offers. However, due to layoffs in the industry, it may be likely that you would be fired within a year after you start working. Which of the following offers seem the most appealing to you?

- ○ Salary of $50,000 with 0% probability of being laid off
- ○ Salary of $60,000 with 15% probability of being laid off
- ○ Salary of $70,000 with 30% probability of being laid off
- ○ Salary of $80,000 with 35% probability of being laid off

UNIVERSITY OF MICHIGAN

You are back on the job market again. You currently already have a job offer that pays $50,000 that you need to respond to within three weeks. You could accept the offer and not have to worry about recruiting for the rest of the year, or you could decline the offer and re-recruit, with no guarantee that you would find something better. If you turn down the offer, there is some chance that you will not have a job when you graduate from college and some chance that you will have a job offer that pays $100,000. In each of the rows below, the probability of getting the higher-paying job varies.

	Accept Job Offer	Re-recruit
A	$50,000 guaranteed	10% chance of $100,000
B	$50,000 guaranteed	20% chance of $100,000
C	$50,000 guaranteed	30% chance of $100,000
D	$50,000 guaranteed	40% chance of $100,000
E	$50,000 guaranteed	50% chance of $100,000
F	$50,000 guaranteed	60% chance of $100,000

Which option would you choose?

- ○ A
- ○ B
- ○ C
- ○ D
- ○ E
- ○ F

Indicate on a scale from 1 to 10 on how likely you are to agree with the following:

Completely disagree									Completely agree	
1	2	3	4	5	6	6	7	8	9	10

College is a risky investment due to student loans

College prepares me for my plans post-graduation

Graduating from college would allow me to make a lot of money in the future

The benefits that come with graduating from college outweigh the costs

You don't need to go to college to be successful

What proportion of your university's students do you believe have a dining plan of some sort?

0 10 20 30 40 50 60 70 80 90 100

Proportion with Dining Plan

Which of the following streaming services do you use?

- [ ] Netflix
- [ ] Disney+
- [ ] Hulu
- [ ] HBO Max
- [ ] Amazon Prime
- [ ] None of the above

Approximately how many novels have you read in 2022?

0 10 20 30 40 50 60 70 80 90 100

Novels

Which of the following fast-food restaurants do you have a positive opinion of?

- [ ] Taco Bell
- [ ] Wendy's
- [ ] McDonald's
- [ ] Burger King
- [ ] Chick-fil-A
- [ ] None of the above

Which of the following retailers do you have a positive opinion of?

- ☐ Amazon
- ☐ Walmart
- ☐ Target
- ☐ Costco
- ☐ None of the above

Which operating software does your phone(s) use?

- ☐ iOS
- ☐ Android
- ☐ None of the above

Appendix D: Experiment Trial #2 Code

```
#------------------------------------
# Econ 497
# Honors Thesis
# By Sharon Ma
# Updated: April 7th, 2023 at 5 PM
#-------------------------------------

#---------------
# Set Directory
#---------------
setwd("~/Downloads")
# install.packages("tidyverse")
library(tidyverse)
# install.packages("stargazer")
library(stargazer)

#--------------
# Import Data
#--------------
orig_survey<-read.csv("orig_survey.csv",encoding="UTF-8")

#---------------
# Create Groups
#---------------

# Subset non/first gen groups

first_gen_group<-orig_survey%>%
  filter(Q34==1) # n=25; matches w var

non_first_gen_group<-orig_survey%>%
  filter(Q34==3) # n=93; matches w var
```

```
# Subset treatment groups

treatment_group<-orig_survey%>%
  filter((Q8==1)|(Q8==2)|(Q8==3)|(Q8==4)|(Q8==5)|(Q8==6)|(Q8==7))# n=60

control_group<-orig_survey%>%
  filter(Q20_1>=0) # n=69

test<-orig_survey%>%
  filter(Q20_1>=0 & Q20_1<=100)

# Subset non/first & treatment groups

treated_first_gen<-orig_survey%>%
  filter((Q8==1)|(Q8==2)|(Q8==3)|(Q8==4)|(Q8==5)|(Q8==6)|(Q8==7))%>%
  filter(Q34==1) # n=14

control_first_gen<-orig_survey%>%
  filter(Q20_1>=0)%>%
  filter(Q34==1) # n=11; numbers add up & matches w var

treated_non_gen<-orig_survey%>%
  filter((Q8==1)|(Q8==2)|(Q8==3)|(Q8==4)|(Q8==5)|(Q8==6)|(Q8==7))%>%
  filter(Q34==3) # n=45

control_non_gen<-orig_survey%>%
  filter(Q20_1>=0)%>%
  filter(Q34==3) # n=48; numbers add up & matches w var
```

```
# Create variables

mod_survey<-orig_survey%>%
  mutate(first_gen_status=
           case_when(
             Q34==1 ~ 1, # 1 if first gen
             Q34==3 ~ 0), # 0 if non first gen
         treatment_status=
           case_when(
             ((Q8==1)|(Q8==2)|(Q8==3)|(Q8==4)|(Q8==5)|(Q8==6)|(Q8==7)) ~ 1, # 1 if treated
             (Q20_1>=0) ~ 0 # 0 if control
           ))

#---------------
# Patience Game
#---------------

# Create variable

mod_survey<-mod_survey%>%
  mutate(patience=
           case_when(
             (Q26_1==1)&(Q26_2==1)&(Q26_3==1)&(Q26_4==1)&(Q26_5==1)~1,
             (Q26_1==1)&(Q26_2==1)&(Q26_3==1)&(Q26_4==1)&(Q26_5==2)~2,
             (Q26_1==1)&(Q26_2==1)&(Q26_3==1)&(Q26_4==2)&(Q26_5==2)~3,
             (Q26_1==1)&(Q26_2==1)&(Q26_3==2)&(Q26_4==2)&(Q26_5==2)~4,
             (Q26_1==1)&(Q26_2==2)&(Q26_3==2)&(Q26_4==2)&(Q26_5==2)~5,
             (Q26_1==2)&(Q26_2==2)&(Q26_3==2)&(Q26_4==2)&(Q26_5==2)~6
           ))


# Summaries

mod_survey%>%
  filter(!is.na(patience) & !is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(patience))

prop.test(c(5.48,5.68),c(93,25))

mod_survey%>%
  filter(!is.na(patience))%>%
  group_by(treatment_status)%>%
  summarize(mean(patience))

prop.test(c(5.57,5.43),c(93,25))

# Regression

patience_reg<-lm(patience~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(patience_reg)
stargazer(patience_reg)
```

```
#-------------
# Risk Games
#-------------

# Variables

mod_survey<-mod_survey%>%
  mutate(risk1=
           case_when(
             Q27==1 ~ 1,
             Q27==2 ~ 2,
             Q27==3 ~ 3,
             Q27==4 ~ 4
           ))

mod_survey<-mod_survey%>%
  mutate(risk2=
           case_when(
             Q28==1 ~ 6,
             Q28==2 ~ 5,
             Q28==3 ~ 4,
             Q28==4 ~ 3,
             Q28==5 ~ 2,
             Q28==6 ~ 1,
           ))

# Summaries

mod_survey%>%
  filter(!is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(risk1))
```

```
prop.test(c(2.83,2.88),c(93,25))

mod_survey%>%
  filter(!is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(risk2))

prop.test(c(1.38,1.8),c(93,25))

# Regression

risk1_reg<-lm(risk1~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(risk1_reg)
stargazer(risk1_reg)

risk2_reg<-lm(risk2~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(risk2_reg)
stargazer(risk2_reg)

# Aggregate

mod_survey%>%
  filter(!is.na(first_gen_status))%>%
  mutate(agg_risk=risk1+risk2)%>%
  group_by(first_gen_status)%>%
  summarize(mean(agg_risk))

prop.test(c(4.20,4.68),c(93,25)) # there may be smth here

temp_mod_survey<-mod_survey%>%
  filter(!is.na(first_gen_status))%>%
  mutate(agg_risk=risk1+risk2)
```

```

agg_risk_reg<-lm(agg_risk~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=temp_mod_survey)
summary(agg_risk_reg)
stargazer(agg_risk_reg)

#-----------------
# Confidence Games
#-----------------

# Summaries

mod_survey%>%
  filter(!is.na(Q29_1),!is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(as.numeric(Q29_1)))

prop.test(c(1.34,1.08),c(93,25))

mod_survey%>%
  filter(!is.na(Q30_1),!is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(as.numeric(Q30_1)))

prop.test(c(0.677,0.52),c(93,25))

# Regressions

confidence1_reg<-lm(Q29_1~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(confidence1_reg)
stargazer(confidence1_reg)


confidence2_reg<-lm(Q30_1~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(confidence2_reg)
stargazer(confidence2_reg)

# Aggregate

mod_survey%>%
  filter(!is.na(Q29_1),!is.na(Q30_1),!is.na(Q31_1),!is.na(first_gen_status))%>%
  mutate(Q29_1=as.numeric(Q29_1),
         Q30_1=as.numeric(Q30_1),
         agg_conf=Q29_1+Q30_1)%>%
  group_by(first_gen_status)%>%
  summarize(mean(agg_conf))

prop.test(c(2.02,1.6),c(93,25))

agg_conf_reg<-lm(agg_conf~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=temp_mod_survey)
summary(agg_conf_reg)
stargazer(agg_conf_reg)

#-----------------
# Competitiveness
#-----------------

mod_survey%>%
  filter(!is.na(Q31_1),!is.na(first_gen_status))%>%
  group_by(first_gen_status)%>%
  summarize(mean(as.numeric(Q31_1)))

prop.test(c(5.78,6.08),c(93,25)) # there may be smth here
```

```
confidence3_reg<-lm(Q31_1~first_gen_status+treatment_status+(first_gen_status*treatment_status),data=mod_survey)
summary(confidence3_reg)
stargazer(confidence3_reg)

#--------
# Exp 1
#--------

# mean first gen 4.090909
# mean non first gen 4.361582

# 44 first gen
# 177 non first gen

prop.test(c(4.090909,4.361582),c(44,177))

prop.test(c(4.435,3.714),c(44,44),alternative="greater")
```

Link to the GitHub: https://github.com/sharonma1218/thesis/blob/main/econ497/econ497code.R