DIVISIONS OF WEALTH: ANALYZING HUMAN PERCEPTION IN RELATION TO ECONOMIC STATUS

Dartmouth College QSS 15

Love Tsai, Angela Zhang

Section I: Introduction

Humans are social creatures who live in groups instead of isolation. The ways in which people interact with and perceive those around them therefore fundamentally affect the cohesion of their communities. Group 6 of QSS 15 is particularly interested in tackling the effects of wealth and in-group biases on human perception. In particular, how does the appearance of wealth influence someone's perception of a peer's friendliness? How does it influence a person's willingness to work with others? Through this study, Group 6 hopes to uncover the relationships between perceived wealth, friendliness, and willingness to cooperate in a group setting.

Notably, while other studies may have occurred in more economically diverse communities, our study takes place in an extremely wealthy community. At Dartmouth College, 69% of students come from the highest-earning 20% of American households. The average income in the Upper Valley is $59,079.16 vs. the College's average family income of $200,400.

Group 6 is interested in measuring how wealth affects human perception in light of these factors: this is where our analysis of in-group bias comes in. We will analyze the effects of wealth on human perception within a community that is inherently already wealthy. Instead of measuring **against** wealth, the way that many other studies do, Group 6's experiment within the College's context will measure for wealth, where being wealthy is the norm instead of the out-group. In other words, our main question is: **how do people of a community view**

the wealthy when such a large proportion of the community is part of the same cohort? Since the College is so largely skewed to the right in terms of annual household income, Group 6 expects that, perhaps contrary to intuition, **there will be no significant bias against students who come from higher-income backgrounds.**

Section II: Design

In order to test the effect of perceived wealth, Group 6 used surveys asking Dartmouth students to rate an actress in a video on:

1. perceived friendliness, and
2. their own willingness to work together in a group

The actress was a non-Dartmouth student so respondents would not respond based on previous interactions.

We had a **Control Group A** and a **Treatment Group B**. Both surveys show a short introduction video of the actress, followed by 3 questions assessing perceived friendliness, willingness to collaborate in a group, and the actress's estimated family income. All variables in the video (duration of speech, background, actress, etc.) were kept constant except for the clothing that was worn: the control included unbranded clothing and the treatment had high-end brands that universally convey wealth among Dartmouth students (such as Canada Goose). Specifics on the script and survey questions appear at the end of this section.

The surveys were sent via email to all undergraduates at Dartmouth College. We were unable to send all emails from one student because Gmail does not allow more than 500 emails to be sent at a time. The group deemed it more important to have all emails be sent at once, regardless of the sender, than to have emails be sent over a 3-day period by one person. This process was randomized to protect against the confounds associated with the variation of the sender.

Distribution of Surveys:

From the full student email list, 9 groups of 482 students were formed using (4338[total # of

students] / 9). Each group was assigned a number between 1-9. Members were assigned their groups using the following R function:

(floor(runif(1, min=1, max=9)))

This function was used until each member had 3 email groups that were non-repeating.

Once subjects clicked the survey link in the email, they were randomly put in Control Group A or Experimental Group B through the (RANDOM FLOW FUNCTION) in Qualtrics. The order of the questions was further randomized through the (RANDOMIZATION FEATURE FUNCTION) in Qualtrics.

We used an **alpha level of .05**, applying a Bonferroni correction to get an **alpha level of .025** to carry out our chi-squared tests.

Survey Script:

The following is the script our actress read for both video A and B:

> "Hi! My name is Emily Smith. I'm a college student from New York who enjoys learning about business and economics. When I'm not in class, you can usually find me watching TV with friends or playing tennis. I also love traveling and meeting new people. I'll see you around!"

Breakdown of Script:

We hoped to eliminate as much potential bias as possible in the script.

1. **Emily** was the most popular baby girl name in 1999.
2. **Smith** is the most common Caucasian last name in the US.
3. **New York City** is in the top 10 of American cities with the largest wealth disparities. People can either be extremely poor or extremely rich, and her living in that area shouldn't tell our survey respondents much about her actual financial status (apart from what they presume based on her appearance).
4. **Business/economics** is the most popular major in the US irrespective of socioeconomic background.

5. We thought that it was important to include a sport/activity. "Watching television with friends" is an activity shared across all income classes. Tennis is a relatively low-barrier entry sport, but people only get good with lots of money put into training, so the dichotomy is applicable for this experiment.

Survey questions:

These three questions were asked in both surveys in random order:

What of the following do you think describes her family's income last year? +less than $10,000 +between $10,000 and $24,999 +between $25,000 and $49,999 +between $50,000 and $74,999 +between $75,000 and $99,999 +between $100,000 and $149,999 +$150,000+

On a scale of 1-10 (1 being least willing to 10 being very willing), how willing are you to collaborate with her on a group project? *

On a scale of 1-10 (1 being not at all to 10 being very much so), how friendly do you think she is?

*Survey A alone was accidentally set to a scale of 1-7 instead of 1-10. Please refer to Section IV: Confounds for further discussion.

Section III: Data Analysis

The Data Analysis Section will be sectioned into 3 parts: 1. Goodness-of-fit test for Randomization, 2. Tests for Willingness to Collaborate, and 3. Tests for Perceived Friendliness.

1. Goodness-of-fit test for Randomization

Table 1: Responses for Groups A and B

	Control Group (A)	Experimental Group (B)	Marginal Response Count
Number of Responses	201	239	440

H_0: Rows and columns are independent

H_1: H_0 is false

$\alpha = 0.05$, degrees of freedom: 1

The Chi-Squared statistic is 3.282.

The p-value is $0.0701 > 0.05$.

We fail to reject H_0 at $\alpha = 0.05$ by the goodness-of-fit test; our control and experimental groups were sampled randomly and we can proceed with the tests.

Note: For the following analyses, responses on the 1-10 scale are grouped into three categories: low < 4, $4 \leqslant$ medium < 8, and $8 \leqslant$ high $\leqslant 10$ because there was an insufficient number of responses to conduct tests with 10 categories.

2. Willingness to Collaborate:

Sample Statistics	**Survey A (Control)**	**Survey B (treatment)**
Sample Size	201	239
Sample Mean	unadjusted = 5.318, adjusted = 7.598 *	7.063
Sample Standard Deviation	unadjusted = 1.435, adjusted = 2.049 *	2.455

The "adjusted" values are to account for Survey A being accidentally set to a scale of 1-7 instead of 1-10. Adjusted value = original value from Survey A times 10/7. Please see Section IV: Confounds for further discussion. For the purposes of comparison, we use the adjusted values in the following sections.

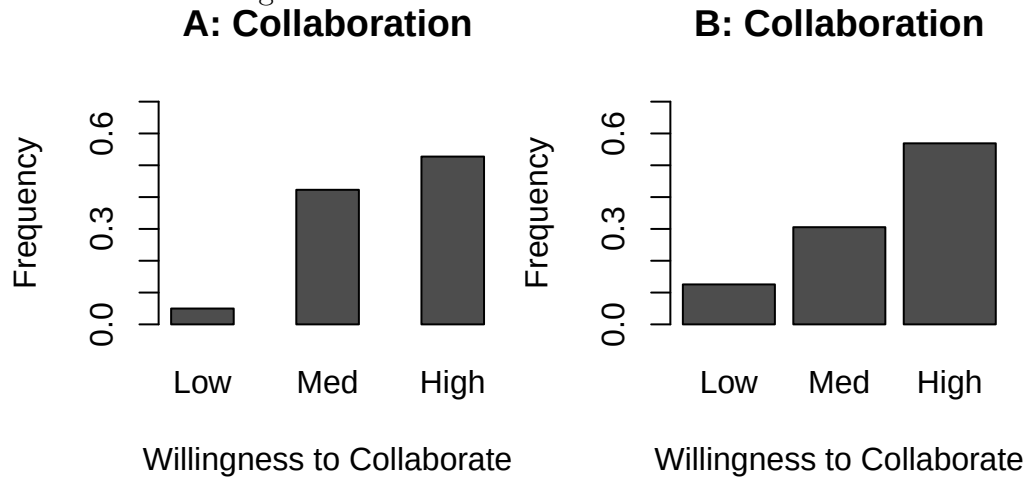

Table 3: Responses for Willingness to Collaborate

	Low	Med	High	Marginal Response Count
Control (A)	10	85	106	201
Experimental (B)	30	73	136	239
Marginal Response Count	40	158	242	440

Chi-Squared Test For Willingness to Collaborate

We conduct a chi-squared test to see whether the appearance of wealth and the willingness to collaborate with the actress are independent. The conditions for running a chi-squared test are met since the expected values for all of our cells in the 3x2 matrix are greater than 1 and the average expected value for all the cells is greater than 5.

H_0: Rows and columns are independent.

H_1: H_0 is false.

$\alpha = 0.05/2 = 0.025$, degrees of freedom: 2

The Chi-Squared statistic is 11.43.

The p-value is $0.00329 < 0.025$.

We have statistical evidence to reject the null hypothesis: the appearance of wealth and the willingness tocollaborate with the actress are not independent.

Table 4: Residuals of chi-squared test

	Low	Med	High
Control (A)	-1.935294	1.509316	-0.4327450
Experimental (B)	1.774786	-1.384138	0.3968543

This table shows that the **"low"** and **"medium"** ratings are contributing most to the chi-squared test statistic. Specifically, less people in the experimental group showed "medium"

willingness to collaborate with the actress than expected (residual = -1.385138 < 0). More people in the experimental group showed "low" willingness to collaborate than (residual = 1.774786 > 0). **This suggests that Dartmouth students are less willing to collaborate with people who appear wealthy.**

Confidence Intervals for Difference in Proportions (A - B)

95% Confidence Interval for Low = -0.0758 +/- 0.0264 = (-0.102, -0.049)

95% Confidence Interval for Medium = 0.117 +/- 0.0458 = (0.0716, 0.163)

95% Confidence Interval for High = -0.0417 +/- 0.0476 = (-0.0893, 0.00593)

These CIs show two things. First, because the interval for High Willingess includes 0, there may be no significant difference between the control and experimental group at $\alpha = 0.05$. Second, because the difference in proportions interval for both Low and Medium Willingess excludes 0, there is some significant difference between the control and experimental group. For Low, we are 95% confident that Control Group A has fewer responses than Experimental Group B. For Medium, we are 95% confident that Control Group A has more responses than Experimental Group B.

Both the residuals and confidence intervals imply the same result. **Dartmouth students are less willing to work with someone who appears extremely wealthy.**

3. Perceived Friendliness:

Sample Statistics	**Survey A (Control)**	**Survey B (treatment)**
Sample Size	201	239
Sample Mean	6.771	6.611
Sample Standard Deviation Perceived Friendliness	2.233	1.85

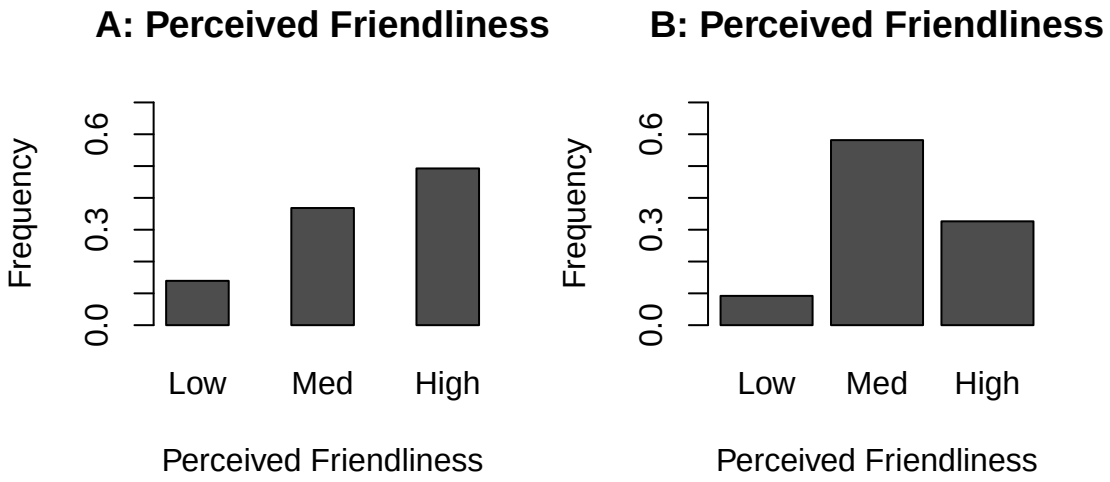

Table 6: Responses

	Low	Med	High	Marginal Response Count
Control (A)	28	74	99	201
Experimental (B)	22	139	78	239
Marginal Response Count	50	213	177	440

Chi-Squared Test for Perceived Friendliness

We conduct a chi-squared test to test whether the appearance of wealth and perceived friendliness are independent. The conditions for running a chi-squared test are met as per the requirements specified earlier.

H_0: Rows and columns are independent.

H_1: H_0 is false.

$\alpha = 0.05/2 = 0.025$, degrees of freedom 2

The Chi-Squared statistic is 19.91.

The p-value is 4.74e-05 < 0.025.

We have statistical evidence to reject the null hypothesis: **the appearance of wealth and perceived friendliness are not independent.**

Table 7: Residuals of chi-squared test

	Low	Med	High
Control (A)	1.0794847	-2.362309	2.017693
Experimental (B)	-0.9899553	2.166386	-1.850351

This table shows that the **"medium"** and **"high"** ratings are contributing the most to the chi-squared test statistic. Specifically, more people in the control group rated "high" level perception of friendliness than expected (residual = 2.017693 > 0). Fewer people in the control group rated a "medium" level of friendliness (residual = -2.362309 < 0), suggesting that Dartmouth students found the control actress friendlier than the wealthy-looking actress (experimental). These two sets of data imply that **Dartmouth students believe those who appear wealthy are less friendly than the average person. Confidence Intervals for Difference in Proportions (A - B) for Perceived Friendliness**

95% Confidence Interval for Low = 0.0473 +/- 0.0308 = (0.0165, 0.078)

95% Confidence Interval for Medium = -0.213 +/- 0.0466 = (-0.26, -0.167)

95% Confidence Interval for High = 0.166 +/- 0.0465 = (0.12, 0.213)

These 95% Confidence Intervals show several things. None of these three intervals include 0. This means that there is significant difference between the control and experimental group at $\alpha = 0.05$ for all three categories. For Low, the entire interval lies above 0, so we are 95% confident that Control Group A would yield greater proportion of responses than Experimental Group B. For Medium, the entire interval lies below 0; we are 95% confident that Control Group A yields a smaller proportion of responses than Experimental Group B. For High, the entire interval lies above 0; we are 95% confident that Control Group A yields a greater proportion of responses than Experimental Group B. **In summary, most people in Control Group A would rate a higher level of friendliness in comparison to Experimental Group B, suggesting that Dartmouth students find wealthier-looking peers less friendly.**

IV. Conclusions: Analysis of Results and Potential Confounds

Our chi-squared tests on the two sets of data found that the probability that wealth has no effect on perceived friendliness was .003, **rejecting** this null hypothesis at $\alpha = .025$. The probability that wealth has no effect on one's willingness to collaborate was 4.74e-05, **rejecting** this null hypothesis at $\alpha = .025$. Therefore, there is statistical evidence to support the theory that wealth is a factor in people's perception of those around them, in both friendliness and willingness to collaborate.

Residuals show that survey takers often rated the actress higher for both willingness to collaborate and perceived friendliness in the control. For willingess to collaborate, more people in the control group gave medium responses than low responses; for perceived friendliness, more people in the control group gave high responses than medium responses. **We conclude that the appearance of wealth may have a negative effect on both willingness to collaborate and perceived friendliness.**

Dartmouth students generally view wealthier-appearing peers with more hostility than less wealthy-appearing peers. It is important to note that the actual differences between control group A and experimental group B did not differ by much ($xbar_{Acollab} = 7.6^*$, $xbar_{Bcollab} = 7.06$; $xbar_{Afriend} = 6.77$, $xbar_{Bfriend} = 6.61$). Even though the data is statistically significant, it is important to question whether a difference between, say, 7.5977257 and 7.0627615 is really significant in practice.

Potential Confounds:

1. 90 emails (out of 4338) bounced back

The surveys that Group 6 used were sent out in emails to Dartmouth's student body. At least 90 of these emails bounced back and were unable to reach the recipient, shrinking our potential sample size. In theory, shrinking the sample size decreases the variability in the data, which would then decrease the chi-squared test statistic and increase our p-value,

making it harder to reject the null. Since we were able to reject the null in both tests, this aspect of the confound isn't as relevant to the discussion. In terms of the confidence intervals, if we hold everything constant but the sample size, our smaller sample size has a wider confidence interval. However, if none of the emails bounced back, our potential sample size could have been bigger and would have resulted in a narrower confidence interval, giving us more precision on the differences between the two sets of data.

It may be important to note that the majority of these emails were aliases (alternate emails that Dartmouth undergraduates can use in lieu of their original ones). The types of people/circumstances that warrant students to create aliases may further confound the data. For example, maybe people create aliases to have shorter emails because they care about being accommodating to others and are generally kinder: had they taken the survey, maybe overall friendliness and willingness to collaborate would have been higher for both groups. There are also ways in which this factor may influence the outcome of the tests, such as if people who create aliases are upperclassmen who have been on campus longer and have worse/better perceptions of people who appear to be wealthy than freshmen. The varying circumstances in which a person may create an alias prevent us from making more concrete claims about how our confidence intervals and tests were affected without running through every single possible scenario.

2. Treatment vs Intent to Treat

It is possible that respondents to our surveys picked up on things in the videos other than the appearance of wealth and created their judgments based off of that. For example, respondents may have viewed Experimental Group B (with the Canada Goose jacket and UGG boots) as flaunting wealth, in which case the overt showcasing of a student's socioeconomic background may more readily lead to a lower score in perceived friendliness. On the other hand, respondents may have seen Control Group A and thought that the actress's casual presentation meant that she didn't care about dressing up more formally for an introduction: they may have rated her lower for willingness to work in a group. Respondents may have also picked up on the backdrop, the script, the speed and delivery of the script, gender, etc.

and made judgments based off of that instead of wealth. This would affect the ideas through which the respondents answer the questions. Then, Group 6's conclusions are unfounded because the experiments do not accurately reflect the effect of the appearance of wealth on human perception of community members.

3. Errors in survey questions: scales of 1 to 10 vs. 1 to 7

As noted above several times, Group 6 had one error in Control Group A, where the scale for group collaboration was corrupted to 1 to 7 instead of 1 to 10. All other questions asking respondents to rate the actress in the video were on a scale from 1 to 10, so in order to have a comparison for this project, we linearly manipulated the results of Control Group A Willingness to Collaborate to be on a scale of 1 to 10, knowing that this is a rudimentary fix at best and the way that people decide ratings is probably not linear. In normal circumstances, we would have redone the surveys and the experiment, but in light of time and resource constraints, Group 6 decided that it was better to manipulate one question's data and extrapolate it to a larger scale instead of manipulating the three other sets of data (even if it is normally better to take data and condense it to a smaller scale). Since we are extrapolating data out into a larger scale, we may have less variance than there should be. Therefore, for our chi-squared test for collaboration, we may have found a smaller test statistic and larger p-value, making it seem like the differences weren't as extreme as they might have been. Also, our confidence interval, keeping everything else constant, may be too narrow: the unaltered sample would have had greater variance, leading to a greater margin of error.

4. Assumption of independent random sample

Because the surveys were taken remotely, we were forced to assume that we had an independent random sample. We cannot tell if friends took the surveys together at the same time, if people overheard us talking about the survey at Novack, or if respondents were otherwise influenced by others while taking the survey. If that were the case, then the sample would not be independent, our actual n is smaller than observed, and we may have had an increase in Type II errors.