# Analysis for First Year Paper

Caleb Griffin

July 2, 2020

Data and Research Design

In order to answer the questions I pose in this paper, and to test the proposed hypotheses, I designed a survey experiment that was distributed to undergraduate political science students at the University of Illinois at Urbana-Champaign. The treatment was designed in the tradition of vignette experiments, specifically, as a paragraph that highlighted Morocco's connections with its European neighbours. The goal of the treatment was to encourage respondents to think of Morocco as a European country, and hence to simulate the category change in regional identity that I introduced earlier. The wording of the treatment is found below:

> "Morocco: A Key Part of Mediterranean Europe:
>
> Scholars tell us that Morocco has been part of the history of southern Europe for thousands of years. Morocco is only 8 miles away from Spain, and there has been movement of people, ideas, and culture back and forth over the course of millennia. Not only are there more than 5 million Moroccans who live in Spain and France, but people from France and Spain also make up the greater part of visitors to Morocco. These countries, along with Morocco, form their own distinct Mediterranean or Southern European region.
>
> In terms of its internal geography, Morocco is a country than spans several climatic regions. This variation is due to its diverse geography. Morocco has two mountain ranges (the Rif and Atlas) which separate the country into different weather zones. To the north, milder conditions prevail and thus it is home to most of Morocco's population. Summers there are warm and winters relatively cool. To the south, the mountains create what geographers call a rain shadow, and thus dryer regions are the norm in the south."

In contrast, a control paragraph was provided to half of participants which only included the second paragraph, which is a description of Morocco's internal geography which gives no explicit information as to its exact location in the world.

The treatment was assigned randomly to participants to break any systematic confounding relationships between my outcome and unobserved variables, and to give a treatment and control group to compare and thus obtain an average treatment effect. Subsequent to receiving the treatment (or not), participants were then asked to locate Morocco on a blank map and draw and label the region which they thought it belonged to. Following that, they were asked to respond to several perceptual questions about Morocco. These were 6 questions that captured different aspects of evaluations one might have about a country, specifically its wealth, economic prospects, level of corruption, and level of ties with the U.S. Respondents were also asked if they would want the U.S. to have stronger ties with Morocco and if they would like to increase immigration from Morocco. Respondents answered using a 5 point ordinal scale ranging from "Strongly Disagree" (which I code as 0) to "Strongly Agree" (which I code as 1). For a couple of questions, the scale was the same although instead of being asked to agree or disagree, they were asked to rate Morocco on the scale in terms of wealth ("Very poor"

(0) to "Very rich" (1)) and current relations "Enemies" (0) to "Allies"(1)). In all cases, however, I scaled each of the six variables to range from 0 to 1, with 0 representing perfectly negative views towards Morocco and 1 representing perfectly positive views towards Morocco. These 6 questions thus represent my dependent variables.

I chose this survey experiment method because it allows me to evaluate whether changing a country's regional category affects perceptions of that country. I can do so by comparing the treatment group's score on each of the dependent variables versus the control group's score. And since assignment to treatment is random in my study, making interpretable comparisons is fairly easy conceptually. Randomization ensures that there is no systematic bias in terms of determining who gets into the treatment and control groups. It breaks all the confounding relations that may exist between variables, both observed and non-observed. As a result, I can say that the difference we are seeing between the treatment and control groups should be the result of exposure to the treatment and not some confounding relationship. The immediate counterargument is that randomization does not ensure balance. I agree with Mutz and Pemantle's approach to this argument (2015). They respond that balance is not necessary for valid inference in experiments. The probability calculation applied automatically makes an allowance for the fact that groups will almost certainly be unbalanced (p.10).

That being said, to satisfy skeptics (and because there is indeed some comfort in knowing that randomization worked as intended), I do actually conduct a balance test made of various indicators. The first is simply the number of respondents who fell into the treatment and control groups. The treatment group had 135 respondents (46 percent of the sample), while the control group has 153 (54 percent of the sample). Ideally the two groups would have an identical number of respondents, but within 8 percentage points of each other is not wildly problematic.[1] The second indicator is comparing the differences in covariate means between the treatment and control groups. These results can be seen in the first four columns of Table 1.

Table 1: Balance Table

| Covariate Mean | Treatment Group | Control Group | Difference | Standardized Difference | Z Score |
|---|---|---|---|---|---|
| Familiarity | 0.89 | 0.88 | 0.01 | 0.02 | 0.16 |
| U.S. Born | 0.91 | 0.8 | 0.11 | 0.31 | 2.6 |
| Ethnocentric | 0.08 | 0.1 | 0.02 | -0.12 | -1.5 |
| White | 0.59 | 0.56 | 0.03 | 0.06 | 0.54 |
| Wealth | 0.66 | 0.68 | 0.02 | -0.06 | -0.5 |
| Female | 0.62 | 0.53 | 0.09 | 0.19 | 1.6 |
| Liberal | 0.71 | 0.67 | 0.05 | 0.18 | 1.5 |
| Christian | 0.48 | 0.47 | 0.01 | 0.02 | 0.18 |
| Overall P.value | 0.33 | | | | |

Despite the potential dangers of randomization in creating imbalance, Table 1 shows remarkable balance in the covariates. All of the listed covariates are measured on a 0 to 1 scale, which means they can typically be converted to proportions or percentages easily. Thus, the largest unstanderdized difference (0.11) in the category of "U.S. Born" means that the treatment group had 11 percentage points more natural-born Americans than the control group. While that one is bordering on becoming worrisome, all of the other categories are smaller than that. In fact a few are nearly perfectly balanced the treatment and control groups, such as level of wealth, Christian affiliation,

[1] The imbalance was created because the survey had to be closed at the end of the semester, when 46 responses were still technically "in progress". By chance, a greater proportion of those respondents happened to be in the treatment group.

and level of Ethnocentrism. Although all the covariates are measured on the same scale, I have also provided a "standardized difference" as another means of comparison. This measure is created by taking the mean of the treatment group minus the mean of the control group for each covariate and divided them by the standard deviation of that covariate.

In terms of the Z score and overall p.value reported, those were created through a type of balance test created by Hansen and Bowers (2008) called "X balance." This is a method that applies Fisher's randomization inference to the question of covariate balance (p.219). The p-value here is testing a two-sided null-hypothesis of no effect. It does so by comparing the differences of means of the regression coefficients to their distributions under hypothetical shuffles of the treatment variable.[2] The Z-scores shown in Table 1 represent where my observed randomization would fall based on a Normal approximation of these shuffled distributions. A high Z-score means that there is a statistically significant likelihood that the observed difference of means between the treatment and control group of a covariate is different from what we would see from perfect randomized shuffling. The only Z-score that is above the conventional 1.96 score that is synonomous with a two-tailed test p.value of smaller than 0.05 is the "U.S. Born" covariate. Randomization has created some imbalance with this variable. The rest are all fine, indeed the overall p.value of 0.33 means that I cannot reject the null hypothesis of no effects–there is no significant difference between the distribution of my randomization and what we would expect given shuffling. This indicates that, on the whole, there are no covariates creating especially serious balance problems between the treatment and control groups. In other words, randomization worked largely as planned, and I have a treatment and control group that are similar enough to facilitate easy comparisons (they differ mostly on whether or not they received random assignment to treatment rather than on the covariates). The only exception is the greater preponderance (by 11 percentage points) of natural-born Americans in the treatment group. Even with this slight imbalance though, the overall balance is good.

While there is undoubtedly artificiality to my method (after all, people are not usually asked to evaluate a country based on several criteria in the real world) it is still part of a rich heritage of research. Survey experiments are a 'tried and true' method of conducting social science research, coming into prominent use as early as 1949 (see Hovland, Lumsdaine, and Sheffield's survey experiment on the effect of war propaganda films on American soldiers in World War II as an example). Part of the longevity of survey experiments must surely be due to their effectiveness. Sniderman (2018) says "survey experiments combine representative samples and randomized assignment–surely a combination that makes for rigorous science" (p.260).[3] The main benefit I see in using a survey experiment is that the treatment effect is clean and easy to measure. It may not mimic the real world exactly, but we can easily see the effect of the treatment. We can also be reasonably confident that every person in the study received the treatment with a survey design. That is, every person will be exposed to the paragraph vignette because it will always appear as part of the survey. There is less worry about some people not receiving treatment. I also think that of the benefits of survey experiments are their familiarity. People are used to taking surveys and therefore take them quite naturally and unaffectedly. If done online with no names collected, people are confident that there responses will not be used to punish them in any way. So survey experiments are a great way to avoid the constant problem of demand effects or social desirability bias.

I am estimating the average treatment effect by comparing the difference between the treatment and control group means of all six outcome variables both seperately and combined. Every outcome variable was measured on a 5 point ordinal scale coded from 0 to 1. By treating this scale as interval

[2]See the R Documentation for Xbalance https://www.rdocumentation.org/packages/RItools/versions/0.1-17/topics/xBalance

[3]It is worth noting that I do not have a representative sample for this study. I see this study as a first step in assessing whether a treatment effect exists, with later studies with representative samples as a goal.

level, I can obtain the average treatment effect by subtracting the mean of the treatment group's dependent variable scores to the mean of the control group. While that can be obtained by a t-test or difference of means test, I chose to use linear regression because I also include covariates in some of the models. Linear regression is a good option in this case because it allows for multivariate analysis while retaining interpretability. The regression coefficient in my models is a type of difference of means because the treatment is binary. That is, we know what the mean of the dependent variable is for the treatment group. It is the intercept plus the regression coefficient. So the regression coefficient of treatment on the outcome is in effect the average treatment effect.

Table 2: Bivariate Regressions

| Outcome Variable | Coefficient | Standard Error | Intercept | Adjusted $R^2$ | F-Statistic |
|---|---|---|---|---|---|
| Wealth | 0.01 | 0.02 | 0.48 | 0 | 0.19 |
| Economic Prospects | 0.01 | 0.02 | 0.56 | 0 | 0.29 |
| Corruption | -0.02 | 0.03 | 0.54 | 0 | 0.59 |
| Immigration | 0.02 | 0.02 | 0.57 | 0 | 0.73 |
| Ties | 0.02 | 0.02 | 0.56 | 0 | 1.17 |
| Current Relations | 0.05** | 0.02 | 0.57 | 0.02 | 5.26 |
| Overall | 0.02* | 0.01 | 0.53 | 0.01 | 2.92 |
| *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$* | | | | | |
| *N = 288* | | | | | |

Table 3: Interaction Coefficients

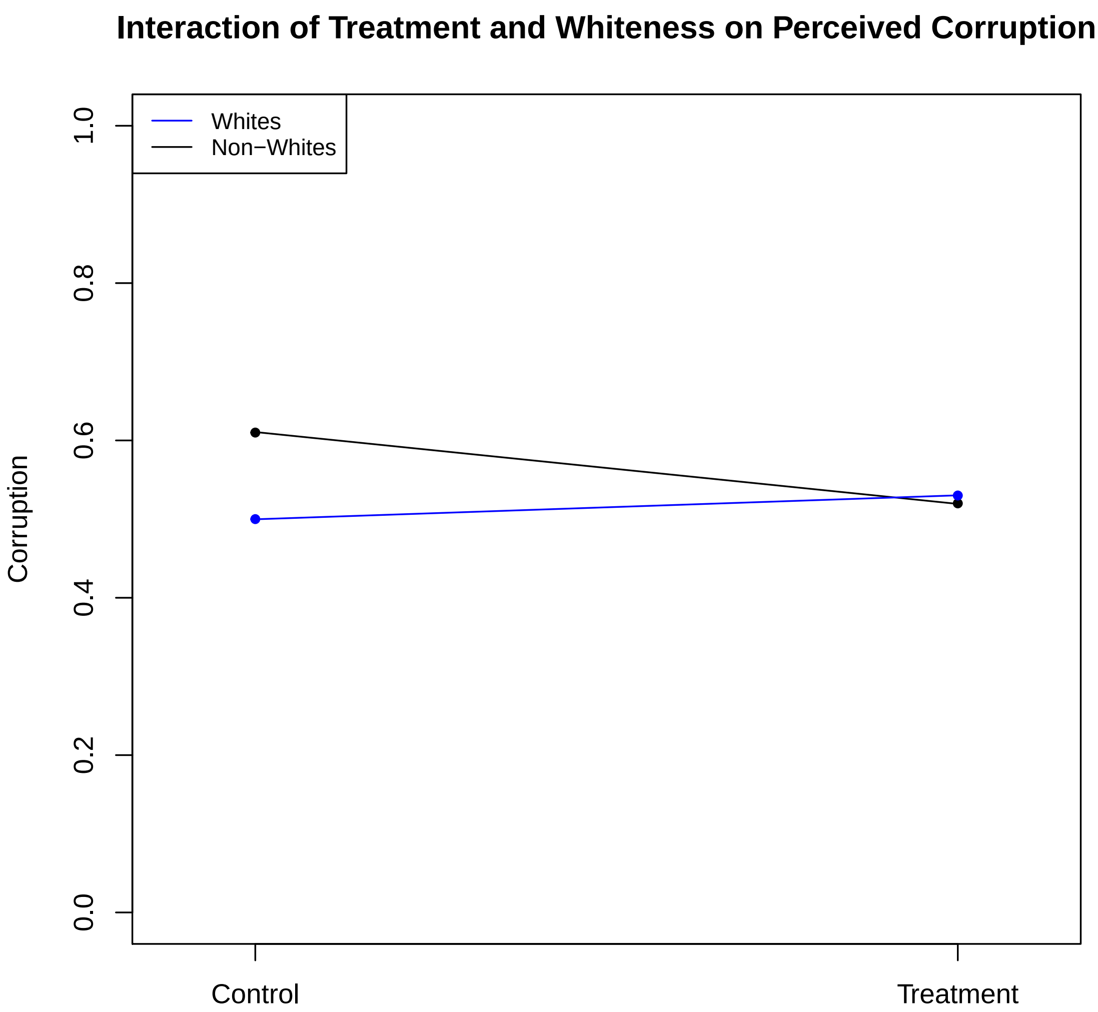
| Outcome Variable | Familiarity | White | Christian | U.S. Born |
|---|---|---|---|---|
| Wealth | -0.04 | -0.01 | 0.03 | 0.02 |
| Economic Prospects | -0.26 | -0.08* | -0.02 | 0.01 |
| Corruption | -0.04 | 0.12** | -0.06 | -0.02 |
| Immigration | -0.001 | -0.07 | -0.02 | 0.005 |
| Ties | -0.02 | 0.004 | 0.06 | -0.09 |
| Current Relations | -0.02 | -0.03 | 0.05 | -0.001 |
| Overall | -0.01 | -0.05** | 0.03 | -0.006 |
| *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$* | | | | |
| *N =* | 287 | 281 | 287 | 287 |

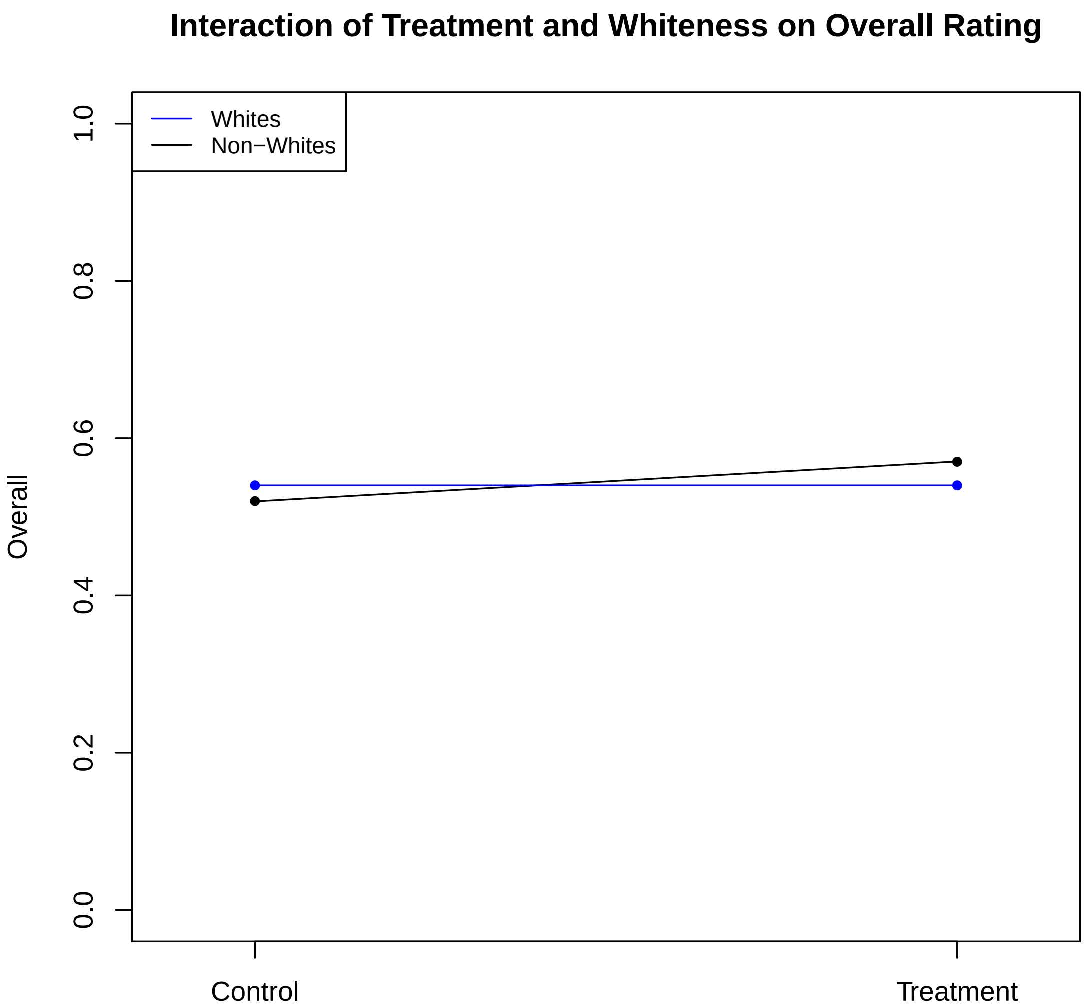
Table 4: Ethnocentric Interaction Coefficients

| Outcome Variable | U.S. Ethnocentric | White Ethnocentric | Christian Ethnocentric |
|---|---|---|---|
| Wealth | -0.06 | -0.06 | -0.33* |
| Economic Prospects | 0.08 | 0.22 | -0.08 |
| Corruption | -0.43** | -0.61** | -0.13 |
| Immigration | 0.06 | 0.2 | -0.16 |
| Ties | -0.009 | 0.04 | -0.09 |
| Current Relations | 0.23 | -0.03 | -0.34* |
| Overall | 0.13 | -0.05** | -0.03 |
| *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$* | | | |
| $N =$ | 218 | 161 | 127 |

```
lm_robust(corrupt ~ treatment*White, data = spdata1)

##                      Estimate Std. Error   t value     Pr(>|t|)    CI Lower
## (Intercept)        0.60727273 0.02798147 21.702677 1.065166e-61  0.55218939
## treatment         -0.08519726 0.04439775 -1.918955 5.601756e-02 -0.17259710
## White             -0.11132035 0.03546559 -3.138827 1.879486e-03 -0.18113667
## treatment:White    0.11937308 0.05471387  2.181770 2.996702e-02  0.01166527
##                      CI Upper  DF
## (Intercept)        0.662356065 277
## treatment          0.002202594 277
## White             -0.041504024 277
## treatment:White    0.227080892 277
```






```
summary(spdata1$Ethnoscorescaled) # mean: 0.09

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.00000 0.00000 0.00125 0.08774 0.12687 0.77500      30

sd(spdata1$Ethnoscorescaled, na.rm = T) # standard deviation: 0.14

## [1] 0.1446464

# 1 standard deviation above mean = 0.23, 1 standard deviation below mean = 0

lm_robust(imm~treatment * Ethnoscorescaled, data = Americans) # original

##                                 Estimate Std. Error     t value     Pr(>|t|)
## (Intercept)                  0.612829702 0.02021333 30.31809902 1.376946e-79
```

```
## treatment                   -0.001082638 0.02872231 -0.03769327 9.699672e-01
## Ethnoscorescaled            -0.231283699 0.16603308 -1.39299769 1.650591e-01
## treatment:Ethnoscorescaled   0.064678142 0.23377220  0.27667166 7.822980e-01
##                                CI Lower   CI Upper  DF
## (Intercept)                  0.57298804 0.65267137 215
## treatment                   -0.05769601 0.05553073 215
## Ethnoscorescaled            -0.55854472 0.09597732 215
## treatment:Ethnoscorescaled  -0.39610069 0.52545697 215

# QUESTION: How do I substitute in the 1-standard deviation above the mean term into the regression
```

Questions

1. I have 6 DVs and 1 combined DV. I think a combined DV may be helpful for ease of interpretation but am not sure if they should be combined. I used Cronbach's Alpha to see the inter-covariance of the items in the scale and it is only 0.45 (unstandardized) and 0.57 (standardized). I am guessing that I should use the raw score since my items are all on a 0-1 scale so there is no need to standardize but I am not sure. In either case, the scores are low, so is that evidence that I should not be making a combined variable?

2. Where is the interaction coefficient in my simple slopes graph?