Gov 1372 - Groups and Identities

Yao Yu

September 23, 2021

Marriage and Partisan Polarization

Iyengar and Westwood (2014) use answers to questions about a child marrying an in-party or out-party spouse as one way of characterizing affective partisan polarization. Some authors have questioned if the way this question is framed too coarsely. In particular, Klar et al. (2018) argue that, by making the prospective child-in-law's partisanship salient, the marriage question may be picking up on respondents dislike of partisanship in general, rather than a dislike of the opposing party.

The in-class survey you took was a partial replication of the Klar et al. (2018) study. We randomized whether you were asked about a prospective child-in-law who "frequently talks about politics," "rarely talks about politics," or a person whose frequency of discussing politics was not mentioned. This last, control, condition matches the wording of the question used in Iyengar and Westwood (2014).

Data Details:

- File Name: `Sep23ClassData_clean.csv`

- Source: These data are from the survey you took in class. The questions are slightly adapted versions of some of the questions used in Klar et al (2018) (see here for the supplemental material of that study with the original questionnaire, if you are interested).

Variable Name	Variable Description
`pid3`	Political party preference
`pid_lean`	If a respondent didn't identify with the Democrats or Republicans in `QID1`, this indicates to which party (or neither) they feel closer
`strongGOP`	Indicator variable for whether the respondent identifies as a Strong Republican
`strongDEM`	Indicator variable for whether the respondent identifies as a Strong Democrat
`strongPARTISAN`	Indicator variable for whether the respondent identifies as a strong member of either major party
`party`	Party variable where those who lean toward either major party are counted as identifying with that party
`treatment`	Which treatment condition the respondent was randomly assigned to
`marryDemocrat`	The respondent's answer to how happy they would be if their child married a Democrat
`marryRepublican`	The respondent's answer to how happy they would be if their child married a Republican
`inPartyHappy`	Indicator variable for whether the respondent would be happy if their child married a member of their own party
`outPartyUnhappy`	Indicator variable for whether the respondent would be unhappy if their child married a member of the other major party

Variable Name	Variable Description
`polarized`	Indicator variable for whether the respondent was affectively polarized

Once again, the .Rmd version of this file has code you can use to load the data.

These data are *not* the raw output from the survey you took. In particular, all of the indicator variables are the result of coding decisions and data processing done by the instructors (based on the procedures used in Klar et al. (2018)). For the first few questions, just open up the data and take a look at it (ask us if you need help viewing the data in spreadsheet format in RStudio).

Question 1

How were the `inPartyHappy` and `outPartyUnhappy` variables defined? Does this seem like a reasonable procedure? Do you notice any missing values? Why are they missing? How might the missing data affect researchers' inferences?

inPartyHappy and outPartyUnhappy are defined based on the following variables: **party**, **marryDemocrat**, **marryRepublican**, **inPartyHappy**, and **outPartyUnhappy**. If the party of the respondent is Democrat and if they responded with "Somewhat happy" or "Very happy" for **marryDemocrat** then **inPartyHappy** would be TRUE. Any other responses for **marryDemocrat** would make **inPartyHappy** FALSE, including Neither happy nor unhappy. This would be flipped for **outPartyUnhappy** and be coded the same way for Republicans using **marryRepublican**. **inPartyHappy** and **outPartyUnhappy** are both NA if the respondent is an Independent or did not respond to the party question. These missing data points could shift any inferences from analysis of this data since it does not take into account the ideology of Independents or those who may not identify with the Democratic or Republican party. Otherwise, I believe that this coding makes sense, but I wonder how coding "Neither happy nor unhappy" responses as NA also might shift the inferences.

Question 2

How was the `polarized` variable defined? Is there another way you might consider coding this variable for individual polarization? What would be an advantage and a disadvantage of your alternative approach?

```
marriage_data %>%
  count(inPartyHappy, outPartyUnhappy, polarized) %>%
  arrange(desc(polarized)) %>%
  gt() %>%
  tab_header(title = "How is polarized calculated?")
```

How is polarized calculated?

inPartyHappy	outPartyUnhappy	polarized	n
TRUE	TRUE	TRUE	39
FALSE	FALSE	FALSE	14
FALSE	TRUE	FALSE	18
TRUE	FALSE	FALSE	8
NA	NA	NA	10

```
marriage_data %>%
  count(strongPARTISAN, inPartyHappy, outPartyUnhappy, polarized) %>%
  arrange(desc(polarized)) %>%
  gt() %>%
  tab_header(title = "How I might calculated polarized")
```

How I might calculated polarized

strongPARTISAN	inPartyHappy	outPartyUnhappy	polarized	n
FALSE	TRUE	TRUE	TRUE	12
TRUE	TRUE	TRUE	TRUE	27
FALSE	FALSE	FALSE	FALSE	12
FALSE	FALSE	TRUE	FALSE	9
FALSE	TRUE	FALSE	FALSE	5
TRUE	FALSE	FALSE	FALSE	2
TRUE	FALSE	TRUE	FALSE	9
TRUE	TRUE	FALSE	FALSE	3
FALSE	NA	NA	NA	10

From this table, we can see that the **polarized** variable is TRUE when **inPartyHappy** and **outPartyUnhappy** is both TRUE. I might code **polarized** to only be TRUE when **strongPARTISAN**, **inPartyHappy**, and **outPartyUnhappy** are all TRUE. Based on the second table, we can see that there are 12 respondents who do not self identify as a strong Democrat or Republican. However, this creates more subjectivity for the **polarized** variable because respondents self-identify if they are strong Democrat or Republican.

Now let's take a look at if there are differences in some of the key outcome variables depending on treatment status. Here is an example of how you might make a graph to look at if the rates of unhappiness with a prospective out-party in-law differ depending on the frequency with which they talk about politics.

```
ggplot(data = marriage_data %>% filter(is.na(outPartyUnhappy)==FALSE)) +
  geom_bar(mapping = aes(x = outPartyUnhappy, y = ..prop.., group = 1), stat = "count") +
  facet_wrap(~treatment) + ylab("Proportion") + xlab("Unhappy with Out-Party Marriage")
```
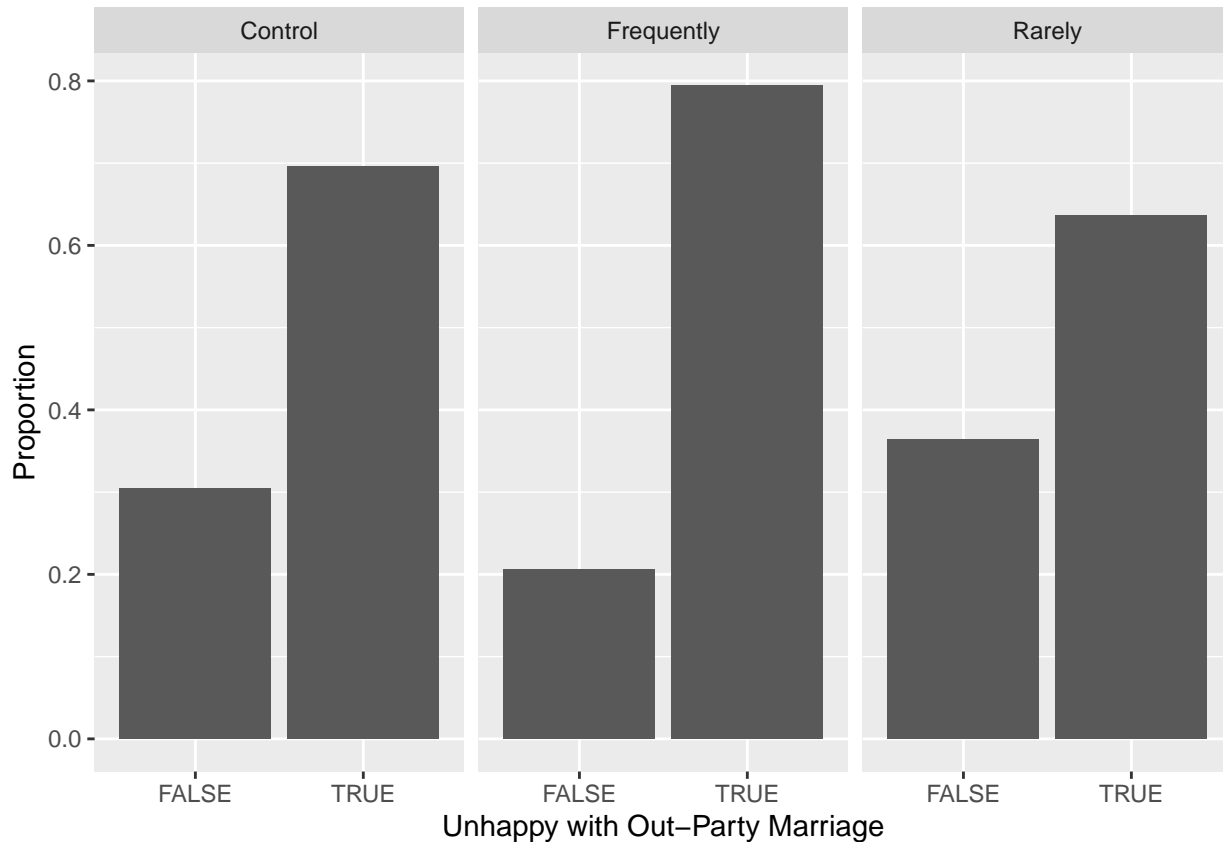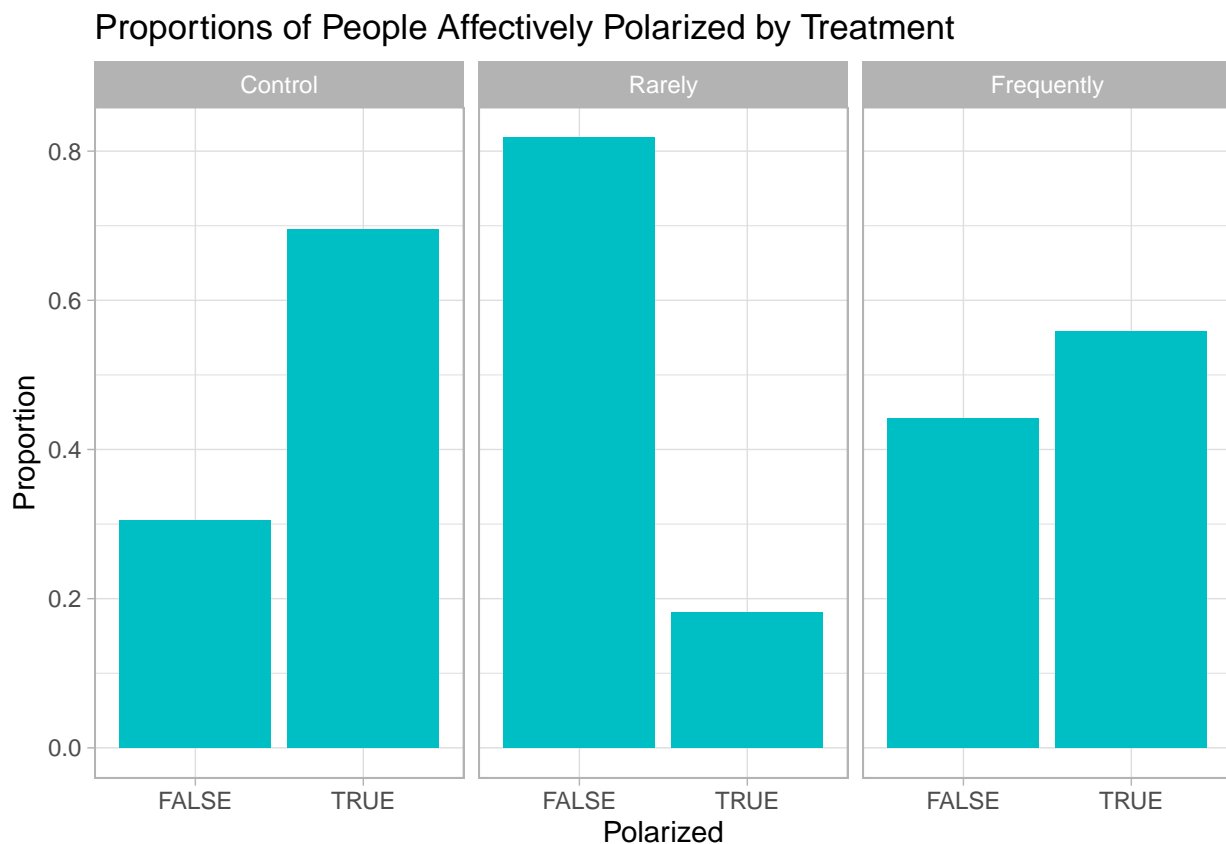

Question 3

Comment on what you see in the example graph. Did the treatment affect unhappiness with a prospective out-party member coming into the family?

From the example graph, we can see that the treatment of whether a prospective out-party member comes into the family might have a small effect on whether a respondent would be unhappy with out-party marriage. If the out-party member frequently discusses politics, then the respondents are on average 10% more unhappy with out-party marriage. On the other side, if the out-party member rarely discusses politics, then the respondents are on average about 7% less unhappy with out-party marriage.

Question 4

Did the different treatment conditions affect the proportions of people who were affectively polarized? Make a plot and explain what you found.

```r
polarized_by_treatment_plot <- marriage_data %>%
  filter(!is.na(polarized)) %>%
  mutate(treatment = fct_relevel(treatment, "Control", "Rarely", "Frequently")) %>%
  ggplot(aes(x = polarized, y = ..prop.., group = 1)) +
  geom_bar(fill = "#00bfc4") +
  facet_wrap(~treatment) +
  theme_light() +
  labs(
    title = "Proportions of People Affectively Polarized by Treatment",
    x = "Polarized",
    y = "Proportion"
  )

polarized_by_treatment_plot
```


Proportions of People Affectively Polarized by Treatment

```r
# png("polarized_by_treatment_plot.png", units="in", width=8, height=5, res=300)
# print(polarized_by_treatment_plot)
# dev.off()
```

Question 5

Take a quick look at Figure 1 and Figure 2 in Klar et al. (2018). How do the results from our in-class data compare to their results? What might explain any differences? If there aren't an notable differences, is there a common pattern across the two datasets that is puzzling? What hypothesis do you have to explain it.

The results from our in-class data follow the general trends from Klar et al. (2018) In Figure 1, the Frequent treatment caused the strongest unhappy with out-party marriage reaction and the rarely treatment caused the weakest unhappy with out-party marriage reaction with the control being somewhere between the two. Data from our class show the same trends although a higher proportion of respondents were unhappy with out-party marriage overall. The data from our class also follow the trends from Figure 2 in Klar et al. (2018) with the same exception that higher proportion of respondents were polarized overall. I think that this higher proportion of unhappiness with out-party marriage and polarization can be attributed to the fact that our sample frame was Harvard undergraduates taking a political class whereas the sample frame in Klar et al. (2018) was Americans generally. Harvard undergraduates who are taking a political course are definitely more likely have strong political views and care more about politics than the general American population. Hence, a higher proportion of respondents would be unhappy with out-party marriage and polarized.

Question 6 (Data Science Question)

We might also be interested in if things looked different for weak vs. strong partisans. Pick one of the two outcome variables you just examined and make a plot that would help us understand if responses within and across treatment groups differ between weak and strong partisans.

```r
# Plotting polarized affects by partisan and treatment
polarized_by_partisan_treatment_plot <- marriage_data %>%
  filter(!is.na(polarized)) %>%
  group_by(strongPARTISAN, treatment) %>%
  summarize(polarized = sum(polarized) / n(),
            .groups = "drop") %>%
  mutate(
    treatment = fct_relevel(treatment, "Control", "Rarely", "Frequently"),
    strongPARTISAN = as.character(strongPARTISAN),
    strongPARTISAN = recode(
      strongPARTISAN,
      "TRUE" = "Strong",
      "FALSE" = "Weak"
    ),
    strongPARTISAN = fct_relevel(strongPARTISAN, "Weak", "Strong")) %>%
  ggplot(aes(x = strongPARTISAN, y = polarized)) +
  geom_col(fill = "#00bfc4", position = "dodge") +
  facet_wrap(~treatment) +
  theme_light() +
  labs(
    title = "Proportions of People Affectively Polarized by Partisan and Treatment",
    x = "Partisan",
    y = "Proportion Polarized"
  )

polarized_by_partisan_treatment_plot
```

Proportions of People Affectively Polarized by Partisan and Treatment
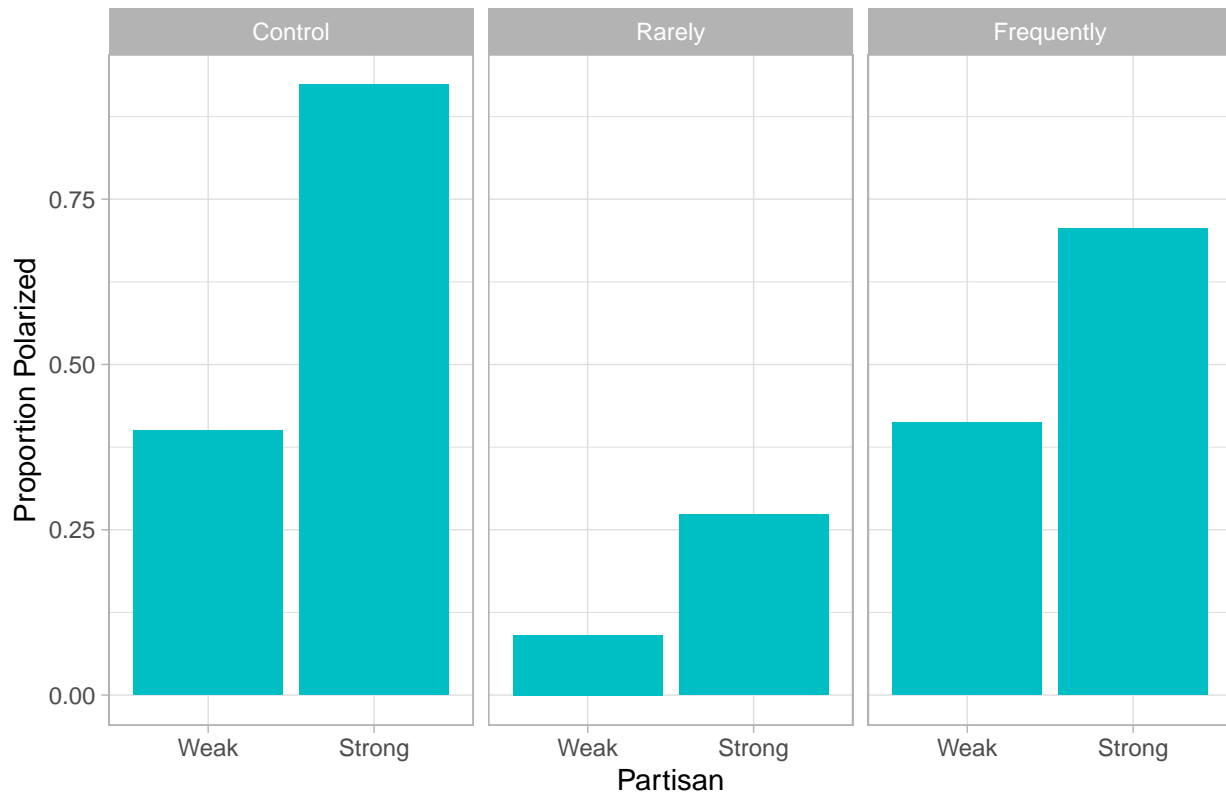

```
# png("polarized_by_partisan_treatment_plot.png", units="in", width=8, height=5, res=300)
# print(polarized_by_partisan_treatment_plot)
# dev.off()
```

Evaluating the polarized effect by treatment and **strongPARTISAN** shows us that the polarized effect is not the same across all levels of partisanship. Strong partisans are much more likely to be polarized than weak partisans according to this figure. Klar et al. (2018) shows the same trends when splitting by partisanship, which makes sense because weak partisans should care less than strong partisans about in-party and out-party marriage.

Question 7

Are there any other issues you can think of that might confound the utility of the marriage question as a measure of affective polarization? If you have any concerns, how might you design a study to evaluate your hypotheses?

One concern that I have about using the marriage question as a measure of affective polarization is that the question relates to the child of an individual and not the individually directly. And since parents are typically very protective of their children, even overly protective sometimes, they just might not want to get into an arugment with their son- or daughter-in-law. I hypothesize that if an alternative question was asked regarding a cousin or distant family member, the measure of affective polarization might be much lower. I would design a survey to replicate Klar et al. (2018) and assign questions about marriage, cousins, and/or other family members randomly to a random sample of Americans.

Question 8

Based on the data and your work on this assignment, are there any changes you would make to the Iyengar and Westwood (2014) study or the Klar et al. (2018) study or extensions of the research that you would like to see? (For example, would you alter the wording of any questions, change the experimental protocol, or come to any different conclusions?)

I would love to extend the Klar et al. (2018) study by stratifying respondents into different regions of America (Northeast, South, West, Midwest) and examine how affective polarization differs across the four regions. I would assume that regions with much more polarization such as the South would have a much higher affective polarization response compared to the other regions. It would also be cool to stratify respondents by state, but the cost of that survey might not be able to justify its usefulness.