Main Replication Code for "Self-Reported Political Ideology"

Eddy S. F. Yeung, Kai Quek

January 18, 2024

```
### Set-up ----
## Clean the R environment and set the working directory
rm(list = ls())
setwd("~/Desktop/ideology-measure/replication")

## Load the required packages
library(tidyverse)   # version 1.3.1
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.2
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to bec
```

```
library(estimatr)    # version 1.0.0
library(bootcorci)   # version 0.0.0.9000
library(cowplot)     # version 1.1.1
```

```
## Warning: package 'cowplot' was built under R version 4.0.2
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
library(texreg)      # version 1.37.5
```

```
## Warning: package 'texreg' was built under R version 4.0.2
```

```
## Version:  1.37.5
```

```
## Date:     2020-06-17
## Author:   Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").
##
## Attaching package: 'texreg'
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(grid)        # version 4.0.1
library(gridExtra)   # version 2.3
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(extrafont)   # version 0.17
```

```
## Registering fonts with R
```

```
library(ltm)         # version 1.2-0
```

```
## Warning: package 'ltm' was built under R version 4.0.5
## Loading required package: MASS
## Warning: package 'MASS' was built under R version 4.0.2
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
## Loading required package: msm
## Warning: package 'msm' was built under R version 4.0.2
## Loading required package: polycor
## Warning: package 'polycor' was built under R version 4.0.5
```

```
library(quanteda)    # version 3.2.0
```

```
## Warning: package 'quanteda' was built under R version 4.0.2
## Package version: 3.2.0
```

```
## Unicode version: 13.0
## ICU version: 69.1
## Parallel computing: 8 of 8 threads used.
## See https://quanteda.io for tutorials and examples.
```

```
library(formattable) # version 0.2.1
```

```
## Warning: package 'formattable' was built under R version 4.0.2
```

```
##
## Attaching package: 'formattable'
##
## The following object is masked from 'package:MASS':
##
##     area
```

```
## Import the dataset
df <- read.csv("survey_data.csv")
nrow(df)
```

```
## [1] 5044
```

```
## Drop non-American respondents
df <- df %>% filter(citizen == 1)
nrow(df)
```

```
## [1] 4357
```

```
## Drop respondents whose survey completion time is less than 5 minutes
median(df$Duration..in.seconds., na.rm = T) / 60
```

```
## [1] 12.56667
```

```
df$more_than_five_mins <- ifelse(df$Duration..in.seconds. >= 5 * 60, "Yes", "No")

# We dropped 528 respondents, 409 of them did not complete the survey
# The sample size is ultimately n = 3828, as reported in main text
table(df$Finished, df$more_than_five_mins)
```

```
##
##       No  Yes
##   0  409  307
##   1  120 3521
```

```
df <- subset(df, more_than_five_mins == "Yes")
nrow(df)
```

```
## [1] 3828
```

```
### Recode sociodemographic variables ----
## Party identification (1 = strong Democrat; 7 = strong Republican)
df <- df %>% mutate(pid = case_when(
  pid1 == 1 & pid2d == 1 ~ 1,
```

```
  pid1 == 1 & pid2d == 2 ~ 2,
  (pid1 == 3 | pid1 == 4) & pid2n == 2 ~ 3,
  (pid1 == 3 | pid1 == 4) & pid2n == 3 ~ 4,
  (pid1 == 3 | pid1 == 4) & pid2n == 1 ~ 5,
  pid1 == 2 & pid2r == 2 ~ 6,
  pid1 == 2 & pid2r == 1 ~ 7))
df$dem <- ifelse(df$pid >= 1 & df$pid <= 3, 1, 0) # (1 = Democrat)
df$gop <- ifelse(df$pid >= 5 & df$pid <= 7, 1, 0) # (1 = Republican)
df$indep <- ifelse(df$pid == 4, 1, 0)             # (1 = independent)

## Race (1 = Black)
df$black <- ifelse(df$racial == 2, 1, 0)

## Gender (1 = female)
df$female <- ifelse(df$gender == 2, 1, 0)

## Education (1 = college graduate)
df$college <- ifelse(df$edu >= 5, 1, 0)

## Political knowledge (0 = least knowledgeable; 4 = most knowledgeable)
df$pol_correct1 <- ifelse(df$know1 == 1, 1, 0)
df$pol_correct2 <- ifelse(df$know2 == 4, 1, 0)
df$pol_correct3 <- ifelse(df$know3 == 2, 1, 0)
df$pol_correct4 <- ifelse(df$know4 == 2, 1, 0)
df$pol_know <- df$pol_correct1 + df$pol_correct2 + df$pol_correct3 + df$pol_correct4
df$sophis <- ifelse(df$pol_know >= 3, 1, 0) # (1 = politically sophisticated)

## Age
df$age <- df$yob + 11

## Income (0 = lowest; 16 = highest)
df$income <- ifelse(df$income == 88, NA, df$income - 1)

### Recode attitudinal variables ----
## Fiscal spending (1 = more services/spending; 7 = cut services/spending)
df$fiscal <- 8 - df$fiscal

## Government responsibility (1 = govt should see to jobs and standard of living;
## 7 = govt should let each person get ahead on own)
df$respon <- df$individual

### Recode ideology variables ----
## Political ideology (1 = extremely liberal; 7 = extremely conservative)
df <- df %>%
  mutate(ideo = case_when(
    randomizer == 1 | randomizer == 2 ~ leftright1,
    randomizer == 3 ~ leftright3,
```

```
    randomizer == 4 ~ leftright4a)) %>%
  mutate(ideo2 = case_when(
    randomizer == 1 | randomizer == 2 ~ leftright1,
    randomizer == 3 ~ leftright3,
    randomizer == 4 ~ leftright4b))
df$liberal <- ifelse(df$ideo2 >= 1 & df$ideo2 <= 3, 1, 0) # (1 = self-id liberal)
df$conserv <- ifelse(df$ideo2 >= 5 & df$ideo2 <= 7, 1, 0) # (1 = self-id conservative)
df$moderate <- ifelse(df$ideo2 == 4, 1, 0)                # (1 = self-id moderate)

## Ideological knowledge (0 = least knowledgeable; 4 = most knowledgeable)
df$ideo_correct1 <- ifelse(df$lib1 == 1, 1, 0)
df$ideo_correct2 <- ifelse(df$lib2 == 2, 1, 0)
df$ideo_correct3 <- ifelse(df$lib3 == 2, 1, 0)
df$ideo_correct5 <- ifelse(df$lib5 == 1, 1, 0)
df$ideo_know <- df$ideo_correct1 + df$ideo_correct2 + df$ideo_correct3 + df$ideo_correct5
table(df$ideo_know)
```

```
##
##   0   1   2   3   4
## 592 496 694 483 509
```

```
## Ideological knowledge after treating "Neither" responses to the second item as correct
## (0 = least knowledgeable; 4 = most knowledgeable)
df$ideo_correct2_alt <- ifelse(df$lib2 == 2 | df$lib2 == 3, 1, 0)
df$ideo_know_2 <- df$ideo_correct1 + df$ideo_correct2_alt + df$ideo_correct3 + df$ideo_correct
table(df$ideo_know_2)
```

```
##
##   0   1   2   3   4
## 358 620 701 514 581
```

```
## Ideological knowledge after including the free trade question
## (0 = least knowledgeable; 5 = most knowledgeable)
df$ideo_correct4 <- ifelse(df$lib4 == 3, 1, 0)
df$ideo_know_alt <- df$ideo_correct1 + df$ideo_correct2 + df$ideo_correct3 +
  df$ideo_correct4 + df$ideo_correct5
table(df$ideo_know_alt)
```

```
##
##   0   1   2   3   4   5
## 401 586 702 487 522  76
```

```
### Define experimental groups ----
df <- df %>% mutate(group = case_when(
  randomizer == 1 | randomizer == 2 ~ 0, # control group
  randomizer == 3 ~ 1,                   # Add Definitions condition
  randomizer == 4 ~ 2))                  # Subtract Labels condition
df$group <- as.factor(df$group)
df0 <- subset(df, group == 0) # create a subset for the control group
```

```r
df1 <- subset(df, group == 1) # create a subset for Add Definitions condition
df2 <- subset(df, group == 2) # create a subset for Subtract Labels condition

### Validate the measure of ideological knowledge ----
## Calculate Cronbach's alpha with the the trade item kept
df_ideo_know <- data.frame(df$ideo_correct1, df$ideo_correct2, df$ideo_correct3,
                           df$ideo_correct4, df$ideo_correct5)
df_ideo_know <- df_ideo_know[complete.cases(df_ideo_know), ]
cronbach.alpha(df_ideo_know) # reported in Appendix E
```

```
##
## Cronbach's alpha for the 'df_ideo_know' data-set
##
## Items: 5
## Sample units: 2774
## alpha: 0.512
```

```r
## Calculate Cronbach's alpha with the free trade item dropped
df_ideo_know_alt <- data.frame(df$ideo_correct1, df$ideo_correct2,
                               df$ideo_correct3, df$ideo_correct5)
df_ideo_know_alt <- df_ideo_know_alt[complete.cases(df_ideo_know_alt), ]
cronbach.alpha(df_ideo_know_alt) # reported in main text
```

```
##
## Cronbach's alpha for the 'df_ideo_know_alt' data-set
##
## Items: 4
## Sample units: 2774
## alpha: 0.652
```

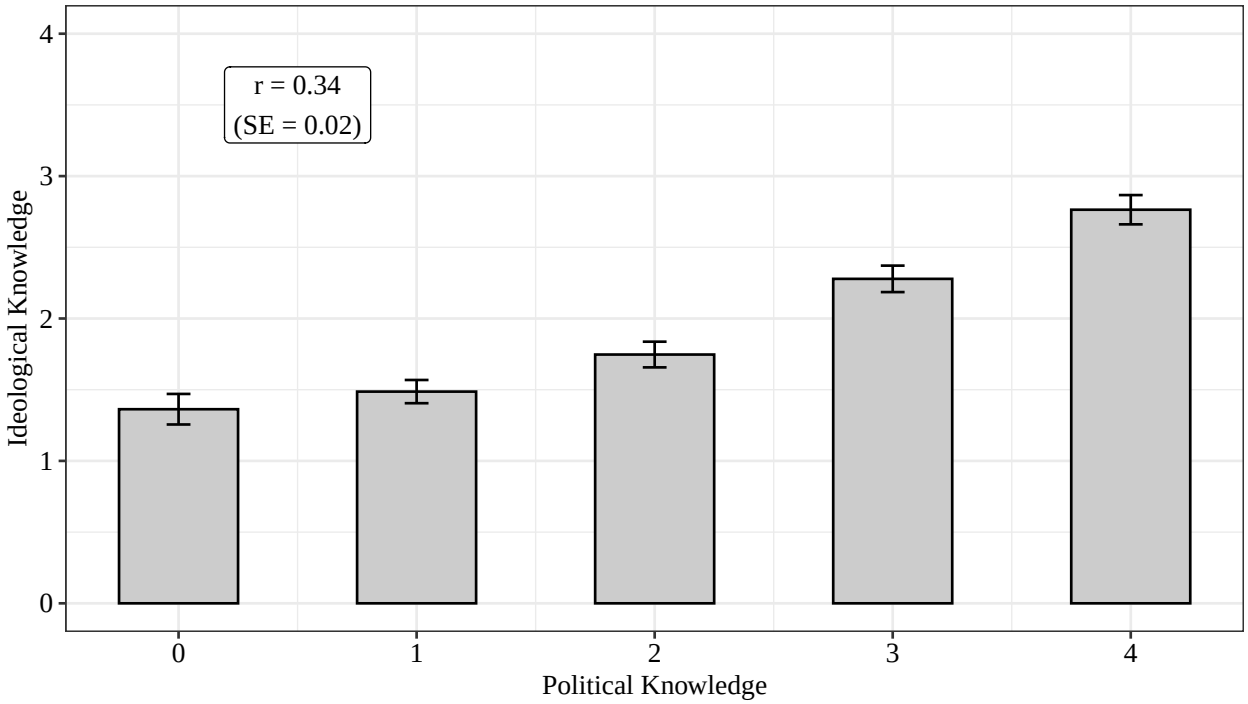
```r
## Correlation between ideological knowledge and political knowledge (with SE)
cor.test.plus <- function(x) {
  list(x, Standard.Error = unname(sqrt((1 - x$estimate^2) / x$parameter)))
}
cor.test.plus(cor.test(df$ideo_know, df$pol_know)) # reported in Appendix F
```

```
## [[1]]
##
##  Pearson's product-moment correlation
##
## data:  df$ideo_know and df$pol_know
## t = 19.139, df = 2755, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3092028 0.3751124
## sample estimates:
##       cor
## 0.3425791
##
```

```
##
## $Standard.Error
## [1] 0.01789909
```

```r
## Visualize how ideological knowledge increases with political knowledge
# Create an empty data frame to store the results first
summary_ideo_know <- data.frame(matrix(NA, nrow = 5, ncol = 4))
colnames(summary_ideo_know) <- c("pol_know", "estimate", "lwr", "upr")
for (i in 1:5) {
  summary_ideo_know[i, 1] <- i - 1
}

# Obtain summary statistics for each level of political knowledge
temp <- df %>%
  group_by(pol_know) %>%
  summarize(mean_ideo_know = mean(ideo_know, na.rm = T),
            sd_ideo_know = sd(ideo_know, na.rm = T),
            n_ideo_know = n()) %>%
  mutate(se_ideo_know = sd_ideo_know / sqrt(n_ideo_know),
         lwr_ci_ideo_know = mean_ideo_know -
           qt(1 - (.05/2), n_ideo_know - 1) * se_ideo_know,
         upr_ci_ideo_know = mean_ideo_know +
           qt(1 - (.05/2), n_ideo_know - 1) * se_ideo_know)
for (i in 1:5) {
  summary_ideo_know[i, c(2:4)] <- c(temp$mean_ideo_know[i],
                                    temp$lwr_ci_ideo_know[i],
                                    temp$upr_ci_ideo_know[i])
}

## Plot the graph
ideo_know_pol_know <-
  ggplot(data = summary_ideo_know, aes(x = pol_know, y = estimate)) +
  geom_bar(stat = "identity", position = position_dodge(),
           color = "black", fill = "grey80", width = .5) +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                linewidth = .5, width = .1,
                position = position_dodge(.9), color = "black") +
  xlab("Political Knowledge") +
  ylab("Ideological Knowledge") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 11, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11)) +
  coord_cartesian(ylim = c(0, 4))
ideo_know_pol_know <- ideo_know_pol_know +
  annotate("label", x = 0.5, y = 3.5,
           family = "Times", label = "r = 0.34\n(SE = 0.02)")
```

```
# Figure S7 in Appendix F
ideo_know_pol_know
```




```
# ggsave(file = "ideo_know_pol_know.pdf", width = 7, height = 4)
```

```
### Table 1: percentage of correct responses to each ideological knowledge question ----
## Create an empty data frame to store the results first
summary_knowledge <- data.frame(matrix(NA, nrow = 5, ncol = 4))
colnames(summary_knowledge) <- c("question", "estimate", "lwr", "upr")
for (i in 1:5) {summary_knowledge[i, 1] <- i}

## Percentage of correct responses to question 1
# (The government should play an active role in supporting social and political change)
temp <- prop.test(x = sum(df$ideo_correct1, na.rm = T),
                  n = sum(!is.na(df$ideo_correct1)),
                  correct = F)
summary_knowledge[1, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 2
# (Social institutions and the free market solve problems better than governments do)
temp <- prop.test(x = sum(df$ideo_correct2, na.rm = T),
                  n = sum(!is.na(df$ideo_correct2)),
                  correct = F)
summary_knowledge[2, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 3
# (A powerful government is a threat to citizens' freedom)
```

```r
temp <- prop.test(x = sum(df$ideo_correct3, na.rm = T),
                  n = sum(!is.na(df$ideo_correct3)),
                  correct = F)
summary_knowledge[3, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 4 (not reported in Table 1; see Appendix E)
# (In general, free trade is good for our country)
temp <- prop.test(x = sum(df$ideo_correct4, na.rm = T),
                  n = sum(!is.na(df$ideo_correct4)),
                  correct = F)
summary_knowledge[4, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 5 (named Q4 but not Q5 in Table 1)
# (The government should play a strong role in the economy and the provision of social service
temp <- prop.test(x = sum(df$ideo_correct1, na.rm = T),
                  n = sum(!is.na(df$ideo_correct5)),
                  correct = F)
summary_knowledge[5, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Show the summary table
# The original Q4 (free trade item) is not reported in Table 1 but reported in Appendix F
# Q5 is renamed as Q4 in Table 1
summary_knowledge
```

```
##   question  estimate       lwr       upr
## 1        1 0.5331174 0.5145321 0.5516111
## 2        2 0.4461649 0.4277636 0.4647150
## 3        3 0.4261527 0.4078714 0.4446382
## 4        4 0.1991358 0.1847028 0.2143999
## 5        5 0.5335014 0.5149099 0.5520004
```

```r
### Table 1: percentage of DKs to each ideological knowledge question ----
## Identify DK responses for each question
df$ideo_DK1 <- ifelse(df$lib1 == 4, 1, 0)
df$ideo_DK2 <- ifelse(df$lib2 == 4, 1, 0)
df$ideo_DK3 <- ifelse(df$lib3 == 4, 1, 0)
df$ideo_DK4 <- ifelse(df$lib4 == 4, 1, 0)
df$ideo_DK5 <- ifelse(df$lib5 == 4, 1, 0)

## Create an empty data frame to store the results first
summary_knowledge_DK <- data.frame(matrix(NA, nrow = 5, ncol = 4))
colnames(summary_knowledge_DK) <- c("question", "estimate", "lwr", "upr")
for (i in 1:5) {summary_knowledge_DK[i, 1] <- i}

## Percentage of DKs for question 1
temp <- prop.test(x = sum(df$ideo_DK1, na.rm = T),
                  n = sum(!is.na(df$ideo_DK1)),
                  correct = F)
```

```r
summary_knowledge_DK[1, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of DKs for question 2
temp <- prop.test(x = sum(df$ideo_DK2, na.rm = T),
                  n = sum(!is.na(df$ideo_DK2)),
                  correct = F)
summary_knowledge_DK[2, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of DKs for question 3
temp <- prop.test(x = sum(df$ideo_DK3, na.rm = T),
                  n = sum(!is.na(df$ideo_DK3)),
                  correct = F)
summary_knowledge_DK[3, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of DKs for question 4
temp <- prop.test(x = sum(df$ideo_DK4, na.rm = T),
                  n = sum(!is.na(df$ideo_DK4)),
                  correct = F)
summary_knowledge_DK[4, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of DKs for question 5
temp <- prop.test(x = sum(df$ideo_DK5, na.rm = T),
                  n = sum(!is.na(df$ideo_DK5)),
                  correct = F)
summary_knowledge_DK[5, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Show the summary table
# The original Q4 (free trade item) is not reported in Table 1
# Q5 is renamed as Q4 in Table 1
summary_knowledge_DK
```

```
##   question  estimate       lwr       upr
## 1        1 0.1177106 0.1062513 0.1302257
## 2        2 0.1498019 0.1370128 0.1635587
## 3        3 0.1340058 0.1218378 0.1471852
## 4        4 0.1584444 0.1453361 0.1724963
## 5        5 0.1152738 0.1039220 0.1276889
```

```r
### Analysis: summary statistics of ideological understanding ----
## Without the free trade question
table(df$ideo_know) # n = 2774 as reported in main text
```

```
##
##   0   1   2   3   4
## 592 496 694 483 509
```

```r
with(df, c(mean(ideo_know, na.rm = T),     # mean = 1.94 as reported in main text
           sd(ideo_know, na.rm = T),       # SD = 1.39 as reported in main text
           median(ideo_know, na.rm = T))) # median = 2 as reported in main text
```

```
## [1] 1.935472 1.391781 2.000000
```

```
## With the free trade question
table(df$ideo_know_alt)
```

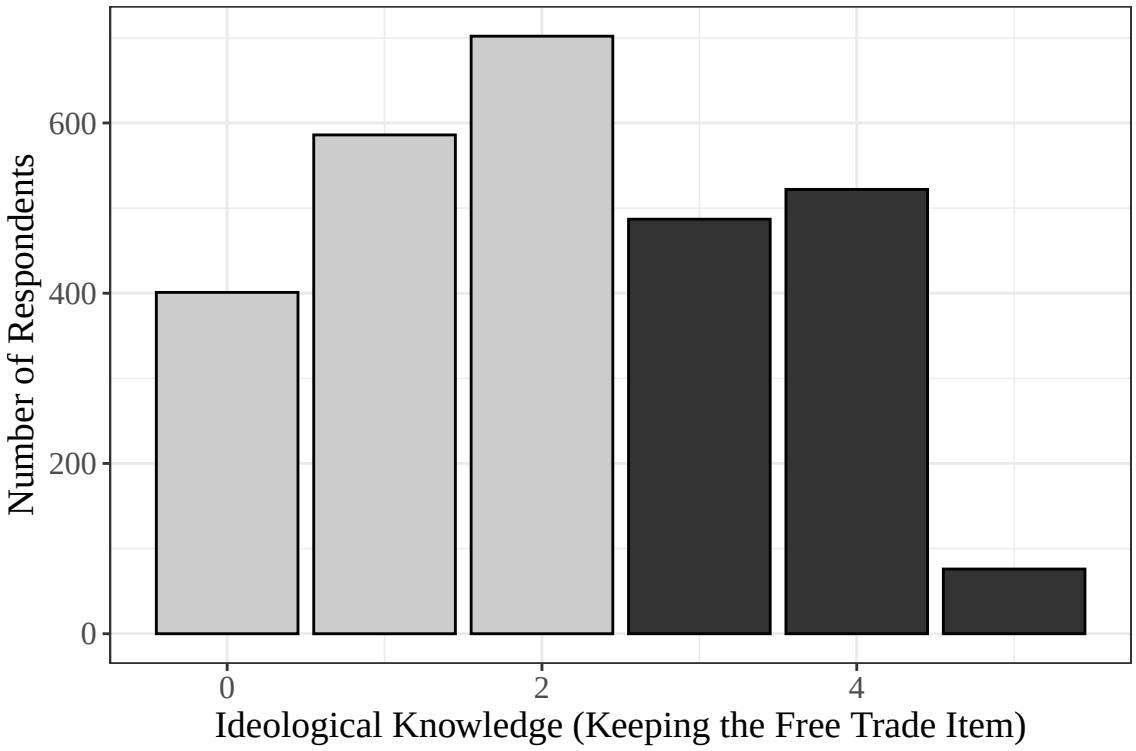
```
##
##   0   1   2   3   4   5
## 401 586 702 487 522  76
```

```
with(df, c(mean(ideo_know_alt, na.rm = T),
           sd(ideo_know_alt, na.rm = T),
           median(ideo_know_alt, na.rm = T)))
```

```
## [1] 2.133742 1.395395 2.000000
```

```
ideo_know_alt <-
  ggplot(df, aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  theme_bw() +
  xlab("Ideological Knowledge (Keeping the Free Trade Item)") +
  ylab("Number of Respondents") +
  theme(text = element_text(family = "Times", size = 14),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")
```

```
# Figure S5 in Appendix E
ideo_know_alt
```




```
# ggsave(file = "ideo_know_alt.pdf", width = 6, height = 4)
```

```
### Analysis: mean number of correct answers by moderates vs. non-moderates ----
temp <- t.test(x = df$ideo_know[df$moderate == 1],
               y = df$ideo_know[df$moderate == 0])
c(temp$estimate, temp$stderr, temp$p.value) # reported in main text
```

```
##    mean of x    mean of y
## 1.302734e+00 2.306461e+00 5.162046e-02 1.363257e-77
```

```
### Analysis: mean number of correct answers by conservatives vs. liberals ----
temp <- t.test(x = df$ideo_know[df$conserv == 1],
               y = df$ideo_know[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in main text
```

```
##    mean of x    mean of y
## 2.4158628081 2.1813725490 0.0614726409 0.0001411859
```

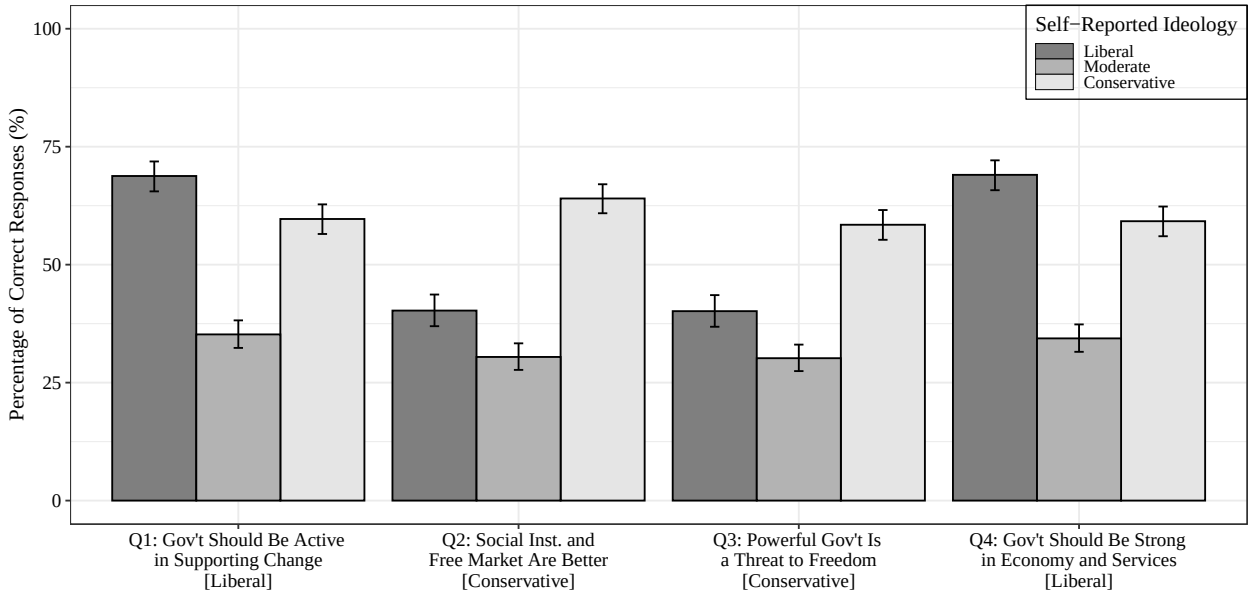
```
### Figure 1: percentage of correct responses to each ideology question by self-reported ideol
## Create an empty data frame to store the results first
summary_know <- data.frame(matrix(NA, nrow = 12, ncol = 5))
colnames(summary_know) <- c("question", "estimate", "lwr", "upr", "Self-Reported Ideology")
summary_know[1:3, 1] <- 1
summary_know[4:6, 1] <- 2
summary_know[7:9, 1] <- 3
summary_know[10:12, 1] <- 5
summary_know[c(1, 4, 7, 10), 5] <- "Liberal"
summary_know[c(2, 5, 8, 11), 5] <- "Conservative"
```

```r
summary_know[c(3, 6, 9, 12), 5] <- "Moderate"

## Percentage of correct responses to question 1 by self-reported ideology
# Liberal
temp <- prop.test(x = sum(df$ideo_correct1[df$liberal == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct1[df$liberal == 1])),
                  correct = F)
summary_know[1, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Conservative
temp <- prop.test(x = sum(df$ideo_correct1[df$conserv == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct1[df$conserv == 1])),
                  correct = F)
summary_know[2, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Moderate
temp <- prop.test(x = sum(df$ideo_correct1[df$moderate == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct1[df$moderate == 1])),
                  correct = F)
summary_know[3, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 2 by self-reported ideology
# Liberal
temp <- prop.test(x = sum(df$ideo_correct2[df$liberal == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2[df$liberal == 1])),
                  correct = F)
summary_know[4, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Conservative
temp <- prop.test(x = sum(df$ideo_correct2[df$conserv == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2[df$conserv == 1])),
                  correct = F)
summary_know[5, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Moderate
temp <- prop.test(x = sum(df$ideo_correct2[df$moderate == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2[df$moderate == 1])),
                  correct = F)
summary_know[6, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 3 by self-reported ideology
# Liberal
temp <- prop.test(x = sum(df$ideo_correct3[df$liberal == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct3[df$liberal == 1])),
                  correct = F)
summary_know[7, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])
```

```r
# Conservative
temp <- prop.test(x = sum(df$ideo_correct3[df$conserv == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct3[df$conserv == 1])),
                  correct = F)
summary_know[8, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Moderate
temp <- prop.test(x = sum(df$ideo_correct3[df$moderate == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct3[df$moderate == 1])),
                  correct = F)
summary_know[9, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Percentage of correct responses to question 5 by self-reported ideology
# Liberal
temp <- prop.test(x = sum(df$ideo_correct5[df$liberal == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct5[df$liberal == 1])),
                  correct = F)
summary_know[10, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Conservative
temp <- prop.test(x = sum(df$ideo_correct5[df$conserv == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct5[df$conserv == 1])),
                  correct = F)
summary_know[11, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

# Moderate
temp <- prop.test(x = sum(df$ideo_correct5[df$moderate == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct5[df$moderate == 1])),
                  correct = F)
summary_know[12, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2])

## Visualize the results
summary_know$`Self-Reported Ideology` <-
  factor(summary_know$`Self-Reported Ideology`,
         levels = c("Liberal", "Moderate", "Conservative"),
         labels = c("Liberal", "Moderate", "Conservative"))
summary_know$question <-
  factor(summary_know$question,
         levels = c(1, 2, 3, 5),
         labels = c("Q1: Gov't Should Be Active\nin Supporting Change\n[Liberal]",
                    "Q2: Social Inst. and\nFree Market Are Better\n[Conservative]",
                    "Q3: Powerful Gov't Is\na Threat to Freedom\n[Conservative]",
                    "Q4: Gov't Should Be Strong\nin Economy and Services\n[Liberal]"))
summary_know$estimate <- summary_know$estimate * 100
summary_know$lwr <- summary_know$lwr * 100
summary_know$upr <- summary_know$upr * 100
ideo_correct <-
```

```r
  ggplot(data = summary_know,
         aes(x = question, y = estimate,
             color = `Self-Reported Ideology`,
             fill = `Self-Reported Ideology`)) +
  geom_bar(stat = "identity", position = position_dodge(.9), color = "black") +
  scale_color_manual(values = c("grey50", "grey70", "grey90")) +
  scale_fill_manual(values = c("grey50", "grey70", "grey90")) +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                linewidth = .5, width = .1,
                position = position_dodge(.9), color = "black") +
  xlab("") +
  ylab("Percentage of Correct Responses (%)") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_blank(),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 100))
```

```r
# Figure 1
ideo_correct
```

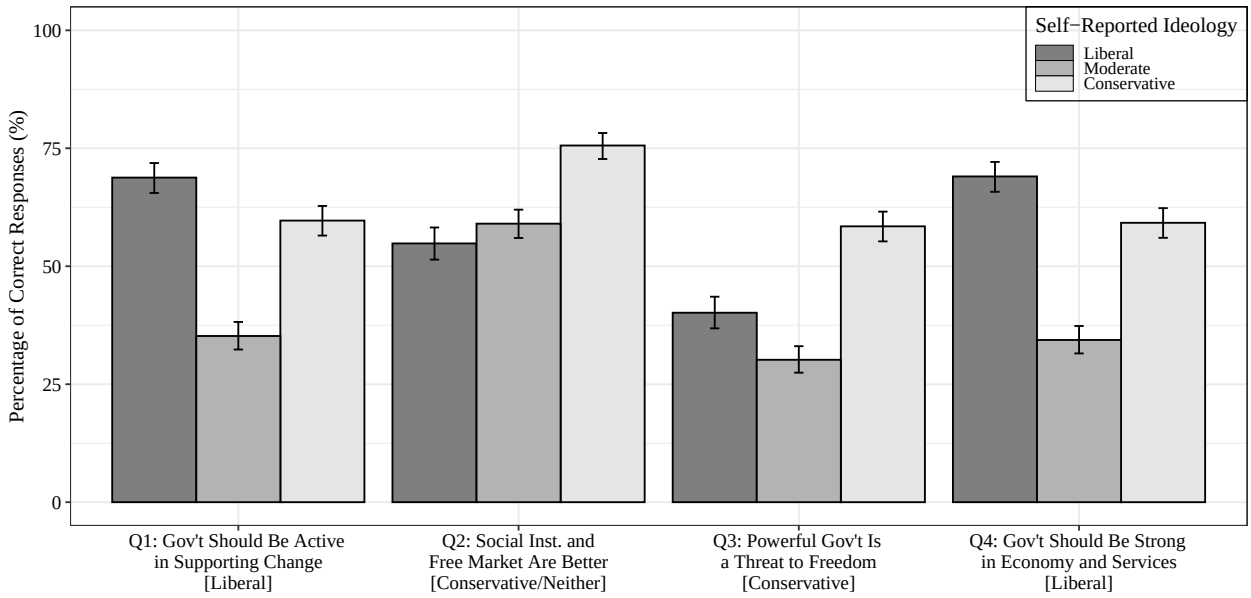



```r
# ggsave(file = "ideo_correct.pdf", width = 10, height = 5)
```

```r
### Alternative analysis for Figure 1: treating "Neither" as correct for question 2 ----
## Percentage of correct responses to question 2 by self-reported ideology
```

```r
# Liberal
temp <- prop.test(x = sum(df$ideo_correct2_alt[df$liberal == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2_alt[df$liberal == 1])),
                  correct = F)
summary_know[4, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2]) * 100

# Conservative
temp <- prop.test(x = sum(df$ideo_correct2_alt[df$conserv == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2_alt[df$conserv == 1])),
                  correct = F)
summary_know[5, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2]) * 100

# Moderate
temp <- prop.test(x = sum(df$ideo_correct2_alt[df$moderate == 1], na.rm = T),
                  n = sum(!is.na(df$ideo_correct2_alt[df$moderate == 1])),
                  correct = F)
summary_know[6, 2:4] <- c(temp$estimate, temp$conf.int[1], temp$conf.int[2]) * 100

summary_know$question <-
  plyr::revalue(summary_know$question,
                c("Q2: Social Inst. and\nFree Market Are Better\n[Conservative]" =
                  "Q2: Social Inst. and\nFree Market Are Better\n[Conservative/Neither]"))
ideo_correct_alt <-
  ggplot(data = summary_know,
         aes(x = question, y = estimate,
             color = `Self-Reported Ideology`,
             fill = `Self-Reported Ideology`)) +
  geom_bar(stat = "identity", position = position_dodge(.9), color = "black") +
  scale_color_manual(values = c("grey50", "grey70", "grey90")) +
  scale_fill_manual(values = c("grey50", "grey70", "grey90")) +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                linewidth = .5, width = .1,
                position = position_dodge(.9), color = "black") +
  xlab("") +
  ylab("Percentage of Correct Responses (%)") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_blank(),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 100))
```

```
# Figure S18 in Appendix F
ideo_correct_alt
```




```
# ggsave(file = "ideo_correct_alt.pdf", width = 10, height = 5)
```

```
### Analysis: knowledge about conservatives and liberals by self-reported ideology ----
## Create a variable indicating a respondent's no. of correct answers to con. questions
df$con_knowledge <- df$ideo_correct2 + df$ideo_correct3

# Alternative operationalization of "correct" answers
df$con_knowledge_alt <- df$ideo_correct2_alt + df$ideo_correct3

## Create a variable indicating a respondent's no. of correct answers to lib. questions
df$lib_knowledge <- df$ideo_correct1 + df$ideo_correct5

## Conservatives' knowledge about con. questions vs. lib. questions
temp <- t.test(x = df$con_knowledge[df$conserv == 1],
               y = df$lib_knowledge[df$conserv == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in main text
```

```
##  mean of x  mean of y
## 1.22483940 1.18843683 0.03864806 0.34636668
```

```
temp <- t.test(x = df$con_knowledge_alt[df$conserv == 1],
               y = df$lib_knowledge[df$conserv == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in Appendix F
```

```
##    mean of x    mean of y
## 1.340471e+00 1.188437e+00 3.736401e-02 4.921961e-05
```

```r
## Liberals' knowledge about con. questions vs. lib. questions
temp <- t.test(x = df$con_knowledge[df$liberal == 1],
               y = df$lib_knowledge[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in Footnote 16
```

```
##    mean of x    mean of y
## 8.041616e-01 1.377451e+00 3.952878e-02 6.372213e-45
```

```r
temp <- t.test(x = df$con_knowledge_alt[df$liberal == 1],
               y = df$lib_knowledge[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in Appendix F
```

```
##    mean of x    mean of y
## 9.498164e-01 1.377451e+00 3.945807e-02 1.777629e-26
```

```r
## Conservatives' knowledge vs. liberals' knowledge about con. questions
temp <- t.test(x = df$con_knowledge[df$conserv == 1],
               y = df$con_knowledge[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in main text
```

```
##    mean of x    mean of y
## 1.224839e+00 8.041616e-01 3.845643e-02 5.639571e-27
```

```r
temp <- t.test(x = df$con_knowledge_alt[df$conserv == 1],
               y = df$con_knowledge_alt[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in Appendix F
```

```
##    mean of x    mean of y
## 1.340471e+00 9.498164e-01 3.709055e-02 3.623573e-25
```

```r
## Conservatives' knowledge vs. liberals' knowledge about lib. questions
temp <- t.test(x = df$lib_knowledge[df$conserv == 1],
               y = df$lib_knowledge[df$liberal == 1])
c(temp$estimate, temp$stderr, temp$p.value) # reported in Footnote 16
```

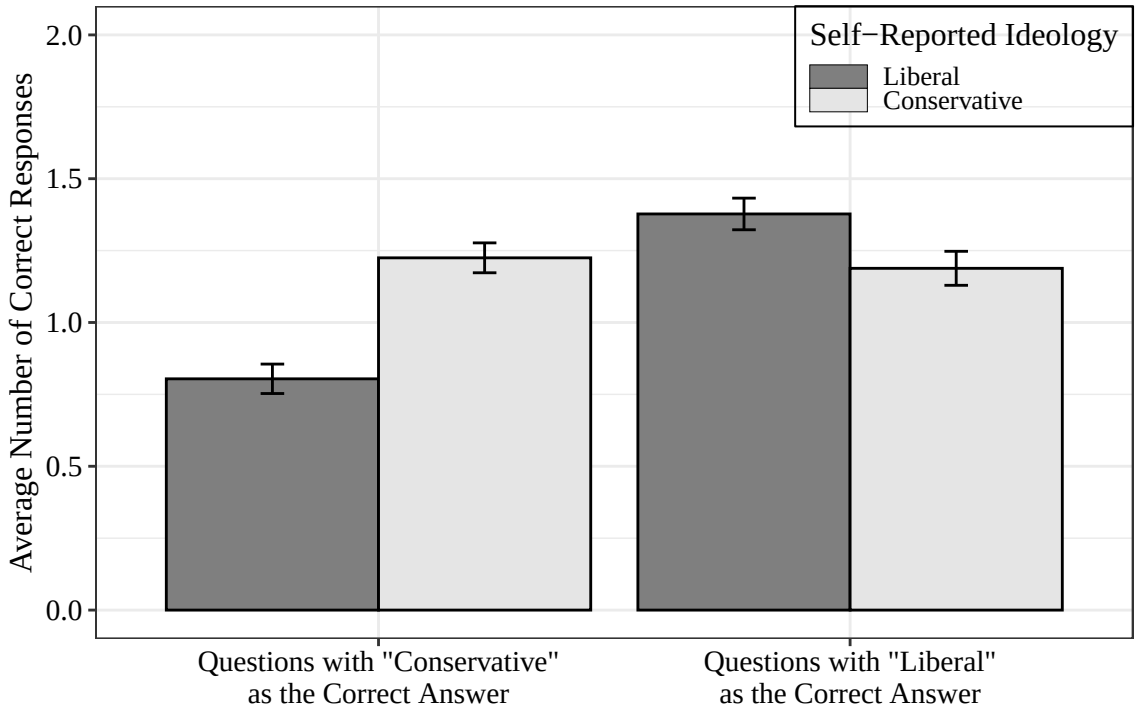
```
##    mean of x    mean of y
## 1.188437e+00 1.377451e+00 3.971523e-02 2.103827e-06
```

```r
### Figure 2: average number of correct responses by question type and self-reported ideology
## Create an empty data frame to store the results first
summary_know2 <- data.frame(matrix(NA, nrow = 4, ncol = 5))
colnames(summary_know2) <- c("group", "estimate", "lwr", "upr", "Self-Reported Ideology")
summary_know2[1:2, 1] <- "Questions with \"Liberal\"\nas the Correct Answer"
summary_know2[3:4, 1] <- "Questions with \"Conservative\"\nas the Correct Answer"
summary_know2[1:4, 5] <- c("Liberal", "Conservative", "Liberal", "Conservative")

## Obtain summary statistics for each group
df$con_lib <- ifelse(df$conserv == 1, 1, ifelse(df$liberal == 1, 0, NA)) # (1 = con.; 0 = lib.)
n_con_know <- sum(!is.na(df$con_knowledge[df$con_lib == 1]))
n_lib_know <- sum(!is.na(df$lib_knowledge[df$con_lib == 0]))
temp <- df %>%
```

```
  group_by(con_lib) %>%
  summarize(mean_con_know = mean(con_knowledge, na.rm = T),
            sd_con_know = sd(con_knowledge, na.rm = T),
            mean_lib_know = mean(lib_knowledge, na.rm = T),
            sd_lib_know = sd(lib_knowledge, na.rm = T)) %>%
  mutate(se_con_know = sd_con_know / sqrt(n_con_know),
         lwr_ci_con_know = mean_con_know -
           qt(1 - (.05/2), n_con_know - 1) * se_con_know,
         upr_ci_con_know = mean_con_know +
           qt(1 - (.05/2), n_con_know - 1) * se_con_know,
         se_lib_know = sd_lib_know / sqrt(n_lib_know),
         lwr_ci_lib_know = mean_lib_know -
           qt(1 - (.05/2), n_lib_know - 1) * se_lib_know,
         upr_ci_lib_know = mean_lib_know +
           qt(1 - (.05/2), n_lib_know - 1) * se_lib_know)
summary_know2[3, c(2:4)] <- c(temp$mean_con_know[1], temp$lwr_ci_con_know[1],
                              temp$upr_ci_con_know[1])
summary_know2[1, c(2:4)] <- c(temp$mean_lib_know[1], temp$lwr_ci_lib_know[1],
                              temp$upr_ci_lib_know[1])
summary_know2[4, c(2:4)] <- c(temp$mean_con_know[2], temp$lwr_ci_con_know[2],
                              temp$upr_ci_con_know[2])
summary_know2[2, c(2:4)] <- c(temp$mean_lib_know[2], temp$lwr_ci_lib_know[2],
                              temp$upr_ci_lib_know[2])

## Plot the graph
summary_know2$`Self-Reported Ideology` <-
  factor(summary_know2$`Self-Reported Ideology`,
         levels = c("Liberal", "Conservative"),
         labels = c("Liberal", "Conservative"))
ideo_con_lib <-
  ggplot(data = summary_know2,
         aes(x = group, y = estimate,
             color = `Self-Reported Ideology`,
             fill = `Self-Reported Ideology`)) +
  geom_bar(stat = "identity", position = position_dodge(.9), color = "black") +
  scale_color_manual(values = c("grey50", "grey90")) +
  scale_fill_manual(values = c("grey50", "grey90")) +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                linewidth = .5, width = .1,
                position = position_dodge(.9), color = "black") +
  xlab("") +
  ylab("Average Number of Correct Responses") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
```

```
        legend.background = element_blank(),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 2))
```

```
# Figure 2
ideo_con_lib
```

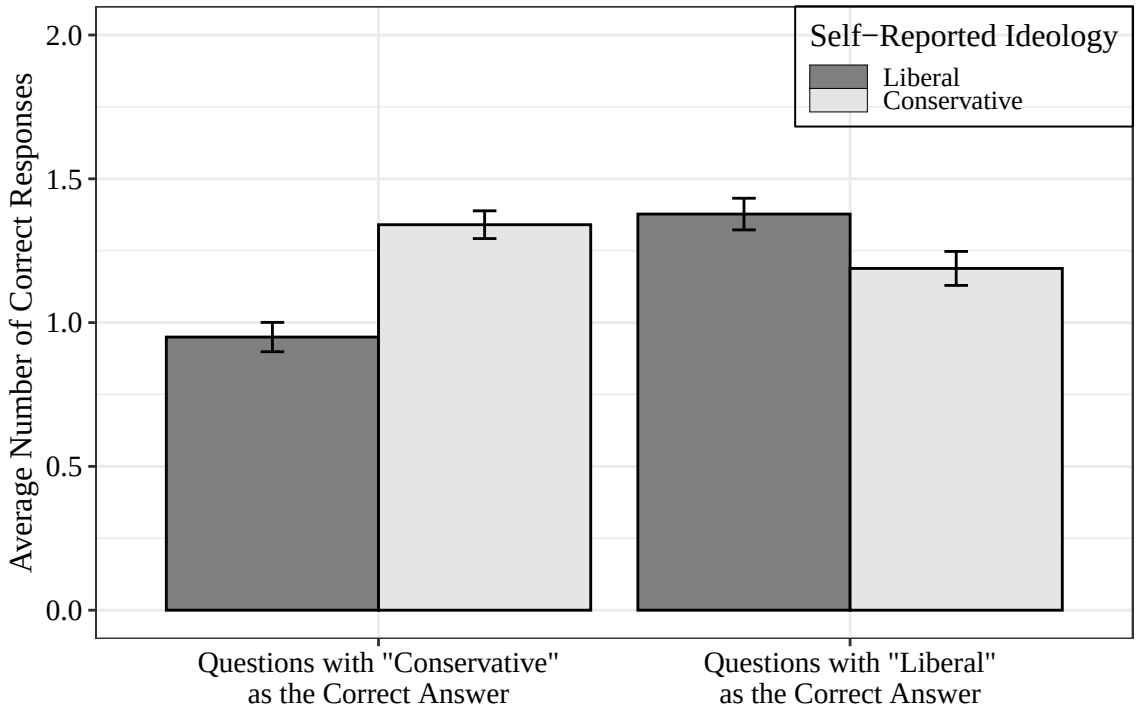



```
# ggsave(file = "ideo_con_lib.pdf", width = 6, height = 4)
```

```
### Alternative analysis for Figure 2: treating "Neither" as correct for question 2 ----
## Obtain summary statistics for each group
n_con_know_alt <- sum(!is.na(df$con_knowledge_alt[df$con_lib == 1]))
temp <- df %>%
  group_by(con_lib) %>%
  summarize(mean_con_know = mean(con_knowledge_alt, na.rm = T),
            sd_con_know = sd(con_knowledge_alt, na.rm = T),
            mean_lib_know = mean(lib_knowledge, na.rm = T),
            sd_lib_know = sd(lib_knowledge, na.rm = T)) %>%
  mutate(se_con_know = sd_con_know / sqrt(n_con_know),
         lwr_ci_con_know = mean_con_know -
           qt(1 - (.05/2), n_con_know_alt - 1) * se_con_know,
         upr_ci_con_know = mean_con_know +
           qt(1 - (.05/2), n_con_know_alt - 1) * se_con_know,
         se_lib_know = sd_lib_know / sqrt(n_lib_know),
         lwr_ci_lib_know = mean_lib_know -
           qt(1 - (.05/2), n_lib_know - 1) * se_lib_know,
```

```
         upr_ci_lib_know = mean_lib_know +
           qt(1 - (.05/2), n_lib_know - 1) * se_lib_know)
summary_know2[3, c(2:4)] <- c(temp$mean_con_know[1], temp$lwr_ci_con_know[1],
                              temp$upr_ci_con_know[1])
summary_know2[1, c(2:4)] <- c(temp$mean_lib_know[1], temp$lwr_ci_lib_know[1],
                              temp$upr_ci_lib_know[1])
summary_know2[4, c(2:4)] <- c(temp$mean_con_know[2], temp$lwr_ci_con_know[2],
                              temp$upr_ci_con_know[2])
summary_know2[2, c(2:4)] <- c(temp$mean_lib_know[2], temp$lwr_ci_lib_know[2],
                              temp$upr_ci_lib_know[2])

## Plot the graph
ideo_con_lib_alt <-
  ggplot(data = summary_know2,
         aes(x = group, y = estimate,
             color = `Self-Reported Ideology`,
             fill = `Self-Reported Ideology`)) +
  geom_bar(stat = "identity", position = position_dodge(.9), color = "black") +
  scale_color_manual(values = c("grey50", "grey90")) +
  scale_fill_manual(values = c("grey50", "grey90")) +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                linewidth = .5, width = .1,
                position = position_dodge(.9), color = "black") +
  xlab("") +
  ylab("Average Number of Correct Responses") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_blank(),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 2))
```

```
# Figure S19 in Appendix F
ideo_con_lib_alt
```




```
# ggsave(file = "ideo_con_lib_alt.pdf", width = 6, height = 4)
```

```
### Analysis: subgroup differences in ideological knowledge ----
## Partisanship
lm_robust(ideo_know ~ gop, data = df,
          subset = (dem == 1 | gop == 1)) # reported in main text
```

```
##               Estimate Std. Error  t value     Pr(>|t|)  CI Lower  CI Upper   DF
## (Intercept) 1.9623699 0.03586689 54.71257 0.000000e+00 1.8920344 2.0327054 2261
## gop         0.3374329 0.05693974  5.92614 3.576704e-09 0.2257733 0.4490925 2261
```

```
## Race
lm_robust(ideo_know ~ black, data = df) # reported in main text
```

```
##                Estimate Std. Error   t value     Pr(>|t|)   CI Lower   CI Upper
## (Intercept)  2.0184874 0.02888231 69.886645 0.000000e+00  1.9618544  2.0751204
## black       -0.5844772 0.06538358 -8.939205 6.995232e-19 -0.7126827 -0.4562718
##                DF
## (Intercept) 2772
## black       2772
```

```
## Sex
lm_robust(ideo_know ~ female, data = df) # reported in main text
```

```
##                Estimate Std. Error   t value     Pr(>|t|)   CI Lower   CI Upper
## (Intercept)  2.2832061 0.03747283  60.92964 0.000000e+00  2.2097286  2.3566836
## female      -0.6588892 0.05145596 -12.80491 1.632774e-36 -0.7597851 -0.5579933
##                DF
## (Intercept) 2772
## female      2772
```

```
## Political knowledge
lm_robust(ideo_know ~ sophis, data = df) # reported in main text
```

```
##              Estimate Std. Error  t value     Pr(>|t|)  CI Lower CI Upper   DF
## (Intercept) 1.5604938 0.03132168 49.82153 0.000000e+00 1.4990775 1.621910 2755
## sophis      0.9153197 0.05182225 17.66268 3.175868e-66 0.8137053 1.016934 2755
```

```
### Figure 3: distribution of ideological knowledge by partisanship, race, sex, and political
## Partisanship
table(df$dem, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 393 254 291 266 321
##   1 199 242 403 217 188
```

```
p1 <- ggplot(subset(df, dem == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Democrats (", italic("n"), " = 1,249)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

table(df$gop, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 438 337 487 266 232
##   1 154 159 207 217 277
```

```
p2 <- ggplot(subset(df, gop == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Republicans (", italic("n"), " = 1,014)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
```

```
        legend.position = "none")

## Race
table(df$black, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 477 416 566 428 493
##   1 115  80 128  55  16
```

```
p3 <- ggplot(subset(df, black == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Blacks (", italic("n"), " = 394)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p4 <- ggplot(subset(df, black == 0), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Non-Blacks (", italic("n"), " = 2,380)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Sex
table(df$female, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 186 179 357 254 334
##   1 406 317 337 229 175
```

```
p5 <- ggplot(subset(df, female == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
```

```
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Female (", italic("n"), " = 1,464)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p6 <- ggplot(subset(df, female == 0), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Male (", italic("n"), " = 1,310)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Political knowledge
table(df$sophis, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 440 342 465 236 137
##   1 146 151 226 244 370
```

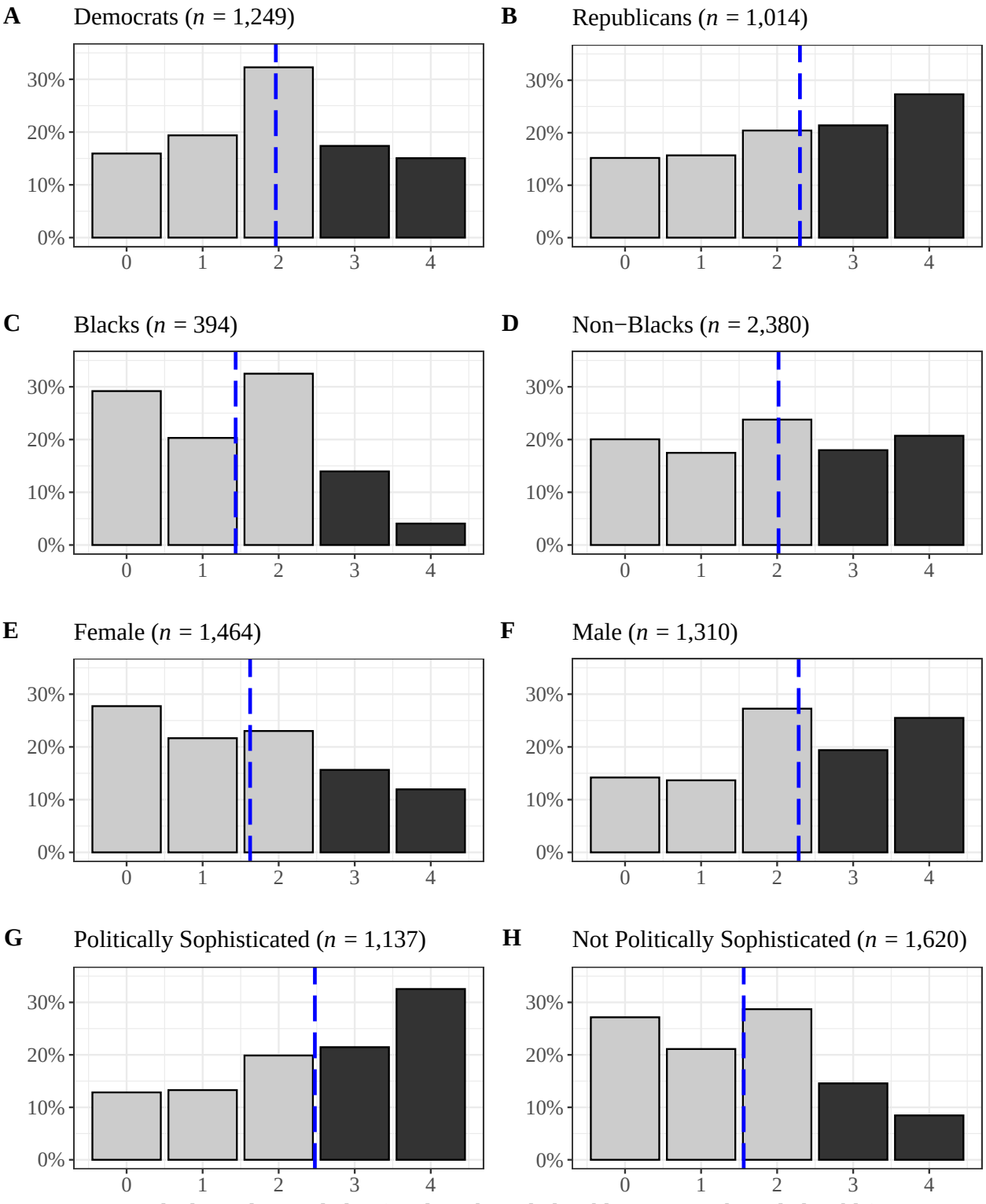
```
p7 <- ggplot(subset(df, sophis == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Politically Sophisticated (", italic("n"), " = 1,137)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p8 <- ggplot(subset(df, sophis == 0), aes(x = ideo_know, fill = ideo_know >= 3)) +
```

```
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Not Politically Sophisticated (", italic("n"), " = 1,620)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 + geom_vline(xintercept = mean(subset(df, dem == 1)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
p2 <- p2 + geom_vline(xintercept = mean(subset(df, gop == 1)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p3 <- p3 + geom_vline(xintercept = mean(subset(df, black == 1)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p4 <- p4 + geom_vline(xintercept = mean(subset(df, black == 0)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p5 <- p5 + geom_vline(xintercept = mean(subset(df, female == 1)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p6 <- p6 + geom_vline(xintercept = mean(subset(df, female == 0)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p7 <- p7 + geom_vline(xintercept = mean(subset(df, sophis == 1)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p8 <- p8 + geom_vline(xintercept = mean(subset(df, sophis == 0)$ideo_know, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)

## Combine into one graph
knowledge_by_group <- plot_grid(p1, p2, p3, p4, p5, p6, p7, p8,
                                labels = "AUTO", ncol = 2,
                                label_fontfamily = "Times")
knowledge_by_group <-
  ggdraw(add_sub(knowledge_by_group,
                 "Ideological Knowledge (0 = least knowledgeable, 4 = most knowledgeable)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```
# Figure 3
knowledge_by_group
```



```r
# ggsave(file = "knowledge_by_group.pdf", width = 8, height = 10)
```

```r
## Conduct two-proportion Z-tests for each pair of subgroups
df$ideo_know_pass <- ifelse(df$ideo_know >= 3, 1, 0)

# All are reported in main text about the subgroup differences in passing rates
prop.test(x = c(sum(df$ideo_know_pass[df$dem == 1], na.rm = T),
                sum(df$ideo_know_pass[df$gop == 1], na.rm = T)),
          n = c(sum(!is.na(df$ideo_know_pass[df$dem == 1])),
                sum(!is.na(df$ideo_know_pass[df$gop == 1]))))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(df$ideo_know_pass[df$dem == 1], na.rm = T), sum(df$ideo_know_pass[df$gop == 1]
## X-squared = 61.36, df = 1, p-value = 4.754e-15
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2040677 -0.1217725
## sample estimates:
##    prop 1    prop 2
## 0.3242594 0.4871795
```

```r
prop.test(x = c(sum(df$ideo_know_pass[df$black == 1], na.rm = T),
                sum(df$ideo_know_pass[df$black == 0], na.rm = T)),
          n = c(sum(!is.na(df$ideo_know_pass[df$black == 1])),
                sum(!is.na(df$ideo_know_pass[df$black == 0]))))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(df$ideo_know_pass[df$black == 1], na.rm = T), sum(df$ideo_know_pass[df$black ==
## X-squared = 62.016, df = 1, p-value = 3.406e-15
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2509504 -0.1625931
## sample estimates:
##    prop 1    prop 2
## 0.1802030 0.3869748
```

```r
prop.test(x = c(sum(df$ideo_know_pass[df$female == 1], na.rm = T),
                sum(df$ideo_know_pass[df$female == 0], na.rm = T)),
          n = c(sum(!is.na(df$ideo_know_pass[df$female == 1])),
                sum(!is.na(df$ideo_know_pass[df$female == 0]))))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(df$ideo_know_pass[df$female == 1], na.rm = T), sum(df$ideo_know_pass[df$female
```

```
## X-squared = 89.216, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2089731 -0.1368243
## sample estimates:
##    prop 1    prop 2
## 0.2759563 0.4488550
```

```
prop.test(x = c(sum(df$ideo_know_pass[df$sophis == 1], na.rm = T),
                sum(df$ideo_know_pass[df$sophis == 0], na.rm = T)),
          n = c(sum(!is.na(df$ideo_know_pass[df$sophis == 1])),
                sum(!is.na(df$ideo_know_pass[df$sophis == 0]))))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(sum(df$ideo_know_pass[df$sophis == 1], na.rm = T), sum(df$ideo_know_pass[df$sophis
## X-squared = 277.59, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2735328 0.3460086
## sample estimates:
##    prop 1    prop 2
## 0.5400176 0.2302469
```

```
### Alternative analysis for Figure 3: treating "Neither" as correct for question 2 ----
## Partisanship
table(df$dem, df$ideo_know_2)
```

```
##
##       0   1   2   3   4
##   0 225 363 303 275 359
##   1 133 257 398 239 222
```

```
p1 <- ggplot(subset(df, dem == 1), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Democrats (", italic("n"), " = 1,249)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

table(df$gop, df$ideo_know_2)
```

```
##
##       0   1   2   3   4
##   0 263 442 483 295 277
##   1  95 178 218 219 304
```

```
p2 <- ggplot(subset(df, gop == 1), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Republicans (", italic("n"), " = 1,014)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Race
table(df$black, df$ideo_know_2)
```

```
##
##       0   1   2   3   4
##   0 294 507 573 451 555
##   1  64 113 128  63  26
```

```
p3 <- ggplot(subset(df, black == 1), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Blacks (", italic("n"), " = 394)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p4 <- ggplot(subset(df, black == 0), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Non-Blacks (", italic("n"), " = 2,380)"))) +
  xlab("") +
```

```
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Sex
table(df$female, df$ideo_know_2)
```

```
##
##       0   1   2   3   4
##   0 104 226 343 272 365
##   1 254 394 358 242 216
```

```
p5 <- ggplot(subset(df, female == 1), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Female (", italic("n"), " = 1,464)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p6 <- ggplot(subset(df, female == 0), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Male (", italic("n"), " = 1,310)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Political knowledge
table(df$sophis, df$ideo_know_2)
```

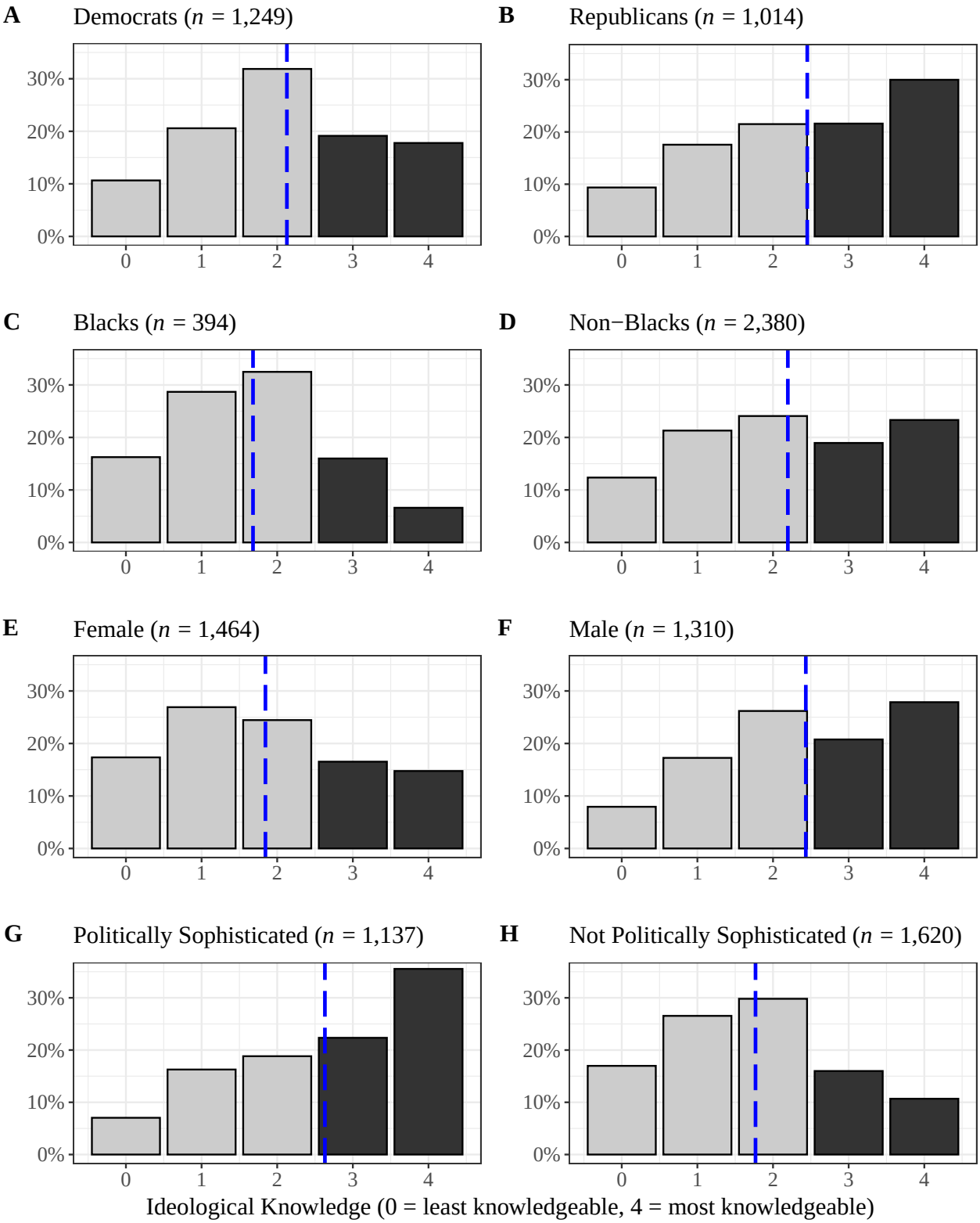
```
##
##       0   1   2   3   4
##   0 275 430 483 259 173
```

```
##   1  80 185 214 254 404
```

```
p7 <- ggplot(subset(df, sophis == 1), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Politically Sophisticated (", italic("n"), " = 1,137)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p8 <- ggplot(subset(df, sophis == 0), aes(x = ideo_know_2, fill = ideo_know_2 >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Not Politically Sophisticated (", italic("n"), " = 1,620)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 + geom_vline(xintercept = mean(subset(df, dem == 1)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p2 <- p2 + geom_vline(xintercept = mean(subset(df, gop == 1)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p3 <- p3 + geom_vline(xintercept = mean(subset(df, black == 1)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p4 <- p4 + geom_vline(xintercept = mean(subset(df, black == 0)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p5 <- p5 + geom_vline(xintercept = mean(subset(df, female == 1)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p6 <- p6 + geom_vline(xintercept = mean(subset(df, female == 0)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p7 <- p7 + geom_vline(xintercept = mean(subset(df, sophis == 1)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p8 <- p8 + geom_vline(xintercept = mean(subset(df, sophis == 0)$ideo_know_2, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
```

```r
## Combine into one graph
knowledge_by_group_alt <- plot_grid(p1, p2, p3, p4, p5, p6, p7, p8,
                                    labels = "AUTO", ncol = 2,
                                    label_fontfamily = "Times")
knowledge_by_group_alt <-
  ggdraw(add_sub(knowledge_by_group_alt,
                 "Ideological Knowledge (0 = least knowledgeable, 4 = most knowledgeable)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```r
# Figure S20 in Appendix F
knowledge_by_group_alt
```




```
# ggsave(file = "knowledge_by_group_alt.pdf", width = 8, height = 10)
```

```
### Alternative analysis for Figure 3: including the free trade item ----
## Partisanship
```

```r
table(df$dem, df$ideo_know_alt)
```

```
##
##       0   1   2   3   4   5
##   0 263 332 306 257 329  38
##   1 138 254 396 230 193  38
```

```r
p1 <- ggplot(subset(df, dem == 1), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Democrats (", italic("n"), " = 1,249)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

table(df$gop, df$ideo_know_alt)
```

```
##
##       0   1   2   3   4   5
##   0 287 415 488 284 241  45
##   1 114 171 214 203 281  31
```

```r
p2 <- ggplot(subset(df, gop == 1), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Republicans (", italic("n"), " = 1,014)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Race
table(df$black, df$ideo_know_alt)
```

```
##
##       0   1   2   3   4   5
##   0 328 479 579 416 503  75
##   1  73 107 123  71  19   1
```

```
p3 <- ggplot(subset(df, black == 1), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Blacks (", italic("n"), " = 394)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p4 <- ggplot(subset(df, black == 0), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Non-Blacks (", italic("n"), " = 2,380)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Sex
table(df$female, df$ideo_know_alt)
```

```
##
##       0   1   2   3   4   5
##   0 119 217 353 230 329  62
##   1 282 369 349 257 193  14
```

```
p5 <- ggplot(subset(df, female == 1), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Female (", italic("n"), " = 1,464)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
```

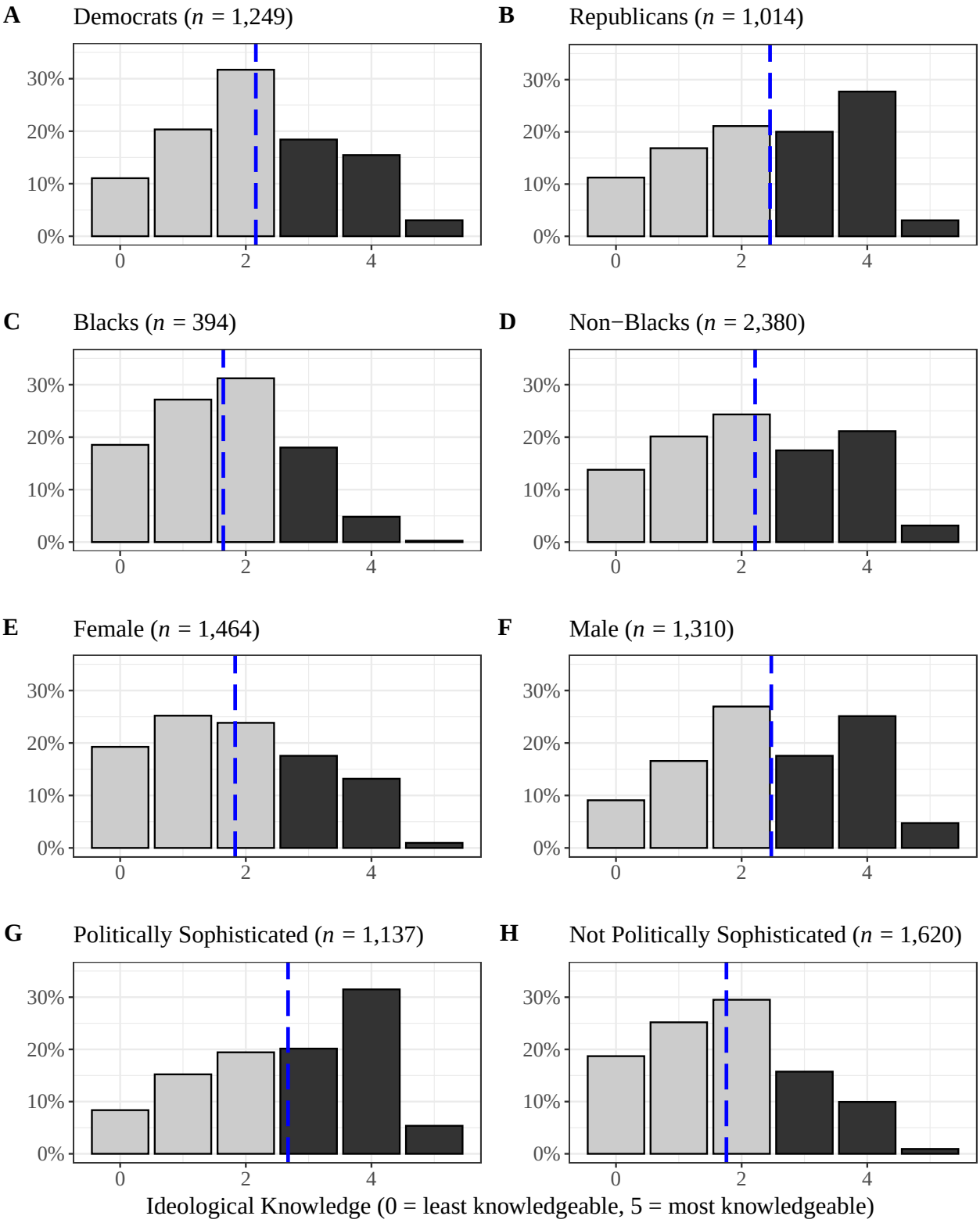
```
        legend.position = "none")

p6 <- ggplot(subset(df, female == 0), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Male (", italic("n"), " = 1,310)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Political knowledge
table(df$sophis, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 440 342 465 236 137
##   1 146 151 226 244 370
```

```
p7 <- ggplot(subset(df, sophis == 1), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Politically Sophisticated (", italic("n"), " = 1,137)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p8 <- ggplot(subset(df, sophis == 0), aes(x = ideo_know_alt, fill = ideo_know_alt >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Not Politically Sophisticated (", italic("n"), " = 1,620)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .35)) +
```

```
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 + geom_vline(xintercept = mean(subset(df, dem == 1)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p2 <- p2 + geom_vline(xintercept = mean(subset(df, gop == 1)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p3 <- p3 + geom_vline(xintercept = mean(subset(df, black == 1)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p4 <- p4 + geom_vline(xintercept = mean(subset(df, black == 0)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p5 <- p5 + geom_vline(xintercept = mean(subset(df, female == 1)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p6 <- p6 + geom_vline(xintercept = mean(subset(df, female == 0)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p7 <- p7 + geom_vline(xintercept = mean(subset(df, sophis == 1)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p8 <- p8 + geom_vline(xintercept = mean(subset(df, sophis == 0)$ideo_know_alt, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)

## Combine into one graph
knowledge_by_group_all <- plot_grid(p1, p2, p3, p4, p5, p6, p7, p8,
                                    labels = "AUTO", ncol = 2,
                                    label_fontfamily = "Times")
knowledge_by_group_all <-
  ggdraw(add_sub(knowledge_by_group_all,
                 "Ideological Knowledge (0 = least knowledgeable, 5 = most knowledgeable)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```
# Figure S6 in Appendix E
knowledge_by_group_all
```




```
# ggsave(file = "knowledge_by_group_all.pdf", width = 8, height = 10)
```

```
### Analysis: DKs across racial and sexual groups ----
## Generate a new variable to indicate the number of DKs for each respondent
```

```r
df$ideo_DK <- df$ideo_DK1 + df$ideo_DK2 + df$ideo_DK3 + df$ideo_DK5

## Racial heterogeneity in DKs
table(df$black, df$ideo_DK)
```

```
##
##        0    1    2    3    4
##   0 1789  279  135   79   98
##   1  273   58   20   15   28
```

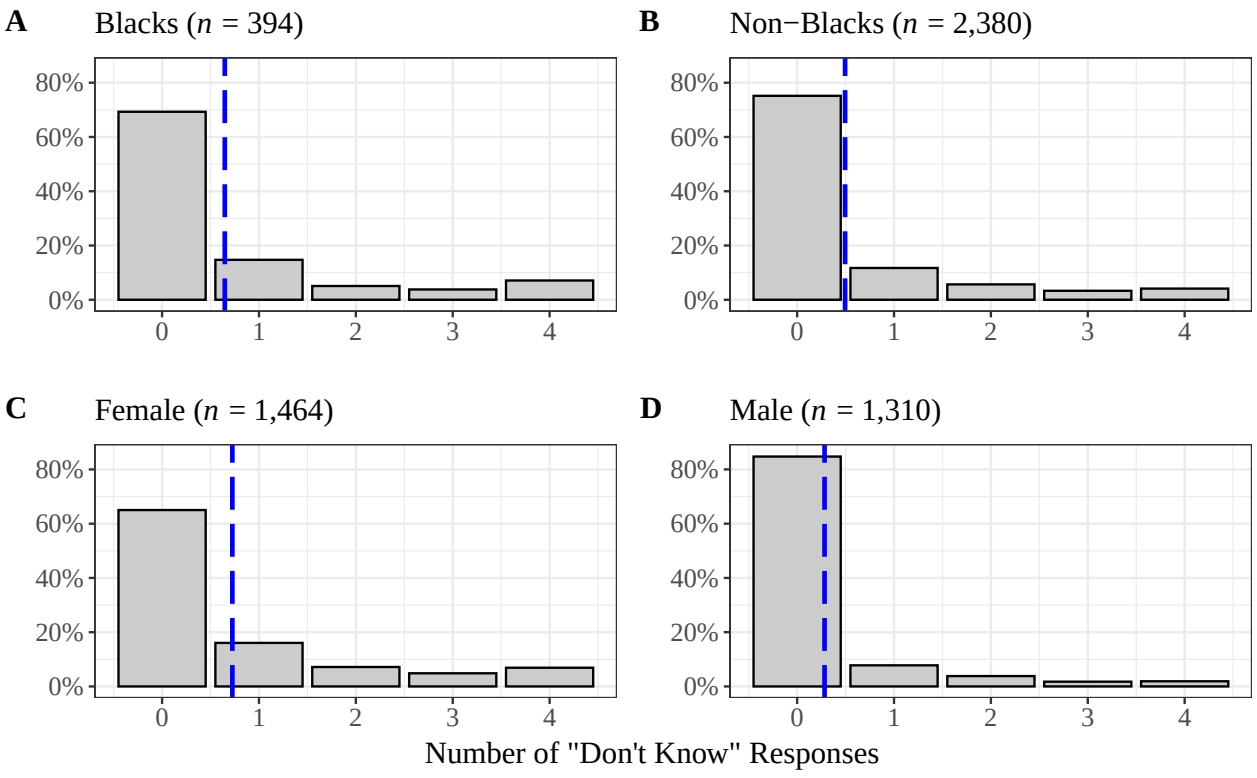
```r
p1 <- ggplot(subset(df, black == 1), aes(x = ideo_DK)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", fill = "grey80", na.rm = T) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Blacks (", italic("n"), " = 394)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .85)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p2 <- ggplot(subset(df, black == 0), aes(x = ideo_DK)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", fill = "grey80", na.rm = T) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Non-Blacks (", italic("n"), " = 2,380)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .85)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Sexual heterogeneity in DKs
table(df$female, df$ideo_DK)
```

```
##
##        0    1    2    3    4
##   0 1110  102   50   23   25
##   1  952  235  105   71  101
```

```r
p3 <- ggplot(subset(df, female == 1), aes(x = ideo_DK)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", fill = "grey80", na.rm = T) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
```

```
  ggtitle(expression(paste("Female (", italic("n"), " = 1,464)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .85)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

p4 <- ggplot(subset(df, female == 0), aes(x = ideo_DK)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", fill = "grey80", na.rm = T) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Male (", italic("n"), " = 1,310)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .85)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 + geom_vline(xintercept = mean(subset(df, black == 1)$ideo_DK, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p2 <- p2 + geom_vline(xintercept = mean(subset(df, black == 0)$ideo_DK, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p3 <- p3 + geom_vline(xintercept = mean(subset(df, female == 1)$ideo_DK, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)
p4 <- p4 + geom_vline(xintercept = mean(subset(df, female == 0)$ideo_DK, na.rm = T),
                      linetype = "longdash", color = "blue", size = 1)

## Combine into one graph
knowledge_by_group_DK <- plot_grid(p1, p2, p3, p4,
                                   labels = "AUTO", ncol = 2,
                                   label_fontfamily = "Times")
knowledge_by_group_DK <-
  ggdraw(add_sub(knowledge_by_group_DK, "Number of \"Don't Know\" Responses",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```
# Figure S21 in Appendix F
knowledge_by_group_DK
```




```
# ggsave(file = "knowledge_by_group_DK.pdf", width = 8, height = 5)
```

```
## Conduct two-tailed t-tests for each pair of subgroups
lm_robust(ideo_DK ~ black, data = df) # reported in Appendix F
```

```
##               Estimate Std. Error   t value      Pr(>|t|)   CI Lower  CI Upper
## (Intercept) 0.4949580 0.02107551 23.484984 2.195200e-111 0.45363270 0.5362833
## black       0.1522501 0.06351920  2.396915  1.659958e-02 0.02770042 0.2767999
##                 DF
## (Intercept) 2772
## black       2772
```

```
lm_robust(ideo_DK ~ female, data = df) # reported in Appendix F
```

```
##               Estimate Std. Error  t value     Pr(>|t|)  CI Lower  CI Upper   DF
## (Intercept) 0.2832061 0.02165229 13.07973 5.593452e-38 0.2407499 0.3256623 2772
## female      0.4422037 0.03831170 11.54226 3.900233e-30 0.3670814 0.5173261 2772
```

```
## Analyze subgroup differences of ideological knowledge by
## assuming all DKs would have been correct
df$ideo_correct1_DK <- ifelse(df$ideo_correct1 == 1 | df$ideo_DK1 == 1, 1, 0)
df$ideo_correct2_DK <- ifelse(df$ideo_correct2 == 1 | df$ideo_DK2 == 1, 1, 0)
df$ideo_correct3_DK <- ifelse(df$ideo_correct3 == 1 | df$ideo_DK3 == 1, 1, 0)
df$ideo_correct5_DK <- ifelse(df$ideo_correct5 == 1 | df$ideo_DK5 == 1, 1, 0)
df$ideo_know_DK_correct <- df$ideo_correct1_DK + df$ideo_correct2_DK +
  df$ideo_correct3_DK + df$ideo_correct5_DK
lm_robust(ideo_know_DK_correct ~ black, data = df) # reported in Appendix F
```

```
##                Estimate Std. Error   t value     Pr(>|t|)   CI Lower   CI Upper
## (Intercept)  2.5134454 0.02586052 97.192370 0.000000e+00  2.4627375  2.5641532
## black       -0.4322271 0.06533431 -6.615622 4.423912e-11 -0.5603359 -0.3041183
##                DF
## (Intercept) 2772
## black       2772
```

```
lm_robust(ideo_know_DK_correct ~ female, data = df) # reported in Appendix F
```

```
##                Estimate Std. Error   t value     Pr(>|t|)   CI Lower   CI Upper
## (Intercept)  2.5664122 0.03377133 75.993826 0.000000e+00  2.5001927  2.6326317
## female      -0.2166854 0.04764327 -4.548081 5.646245e-06 -0.3101053 -0.1232656
##                DF
## (Intercept) 2772
## female      2772
```

```
### Analysis: self-reported ideology by experimental condition ----
## Group means by experimental condition
# Control group
lm_robust(ideo ~ 1, data = df0) # reported in main text
```

```
##             Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept) 4.116835 0.03833654 107.3867        0 4.041648 4.192021 1882
```

```
# Add Definitions condition
lm_robust(ideo ~ 1, data = df1) # reported in main text
```

```
##             Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept) 4.024892 0.05901639 68.19956        0  3.90907 4.140714 923
```

```
# Subtract Labels condition
lm_robust(ideo ~ 1, data = df2) # reported in main text
```

```
##             Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)  4.07766 0.05407856 75.40252        0 3.971531 4.183788 939
```

```
## Mean ideology score in Add Definitions condition vs. control
t.test(x = df0$ideo, y = df1$ideo) # statistically insignificant at the 10% level
```

```
##
##  Welch Two Sample t-test
##
## data:  df0$ideo and df1$ideo
## t = 1.3065, df = 1716.4, p-value = 0.1916
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.04608651  0.22997263
## sample estimates:
## mean of x mean of y
##  4.116835  4.024892
```

```r
## Mean ideology score in Subtract Labels condition vs. control
t.test(x = df0$ideo, y = df2$ideo) # statistically insignificant at the 10% level
```

```
##
##  Welch Two Sample t-test
##
## data:  df0$ideo and df2$ideo
## t = 0.59098, df = 1882.7, p-value = 0.5546
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.09083162  0.16918214
## sample estimates:
## mean of x mean of y
##  4.116835  4.077660
```

```r
## Mean ideology score in Add Definitions condition vs. Subtract Labels condition
t.test(x = df1$ideo, y = df2$ideo) # statistically insignificant at the 10% level
```

```
##
##  Welch Two Sample t-test
##
## data:  df1$ideo and df2$ideo
## t = -0.65922, df = 1845.1, p-value = 0.5098
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2097588  0.1042232
## sample estimates:
## mean of x mean of y
##  4.024892  4.077660
```

```r
## Distribution of ideology in Add Definitions condition vs. control
ks.test(x = df0$ideo, y = df1$ideo) # reported in Appendix C
```

```
## Warning in ks.test(x = df0$ideo, y = df1$ideo): p-value will be approximate in
## the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  df0$ideo and df1$ideo
## D = 0.060566, p-value = 0.02119
## alternative hypothesis: two-sided
```

```r
## Distribution of ideology in Subtract Labels condition vs. control
ks.test(x = df0$ideo, y = df2$ideo) # reported in Appendix C
```

```
## Warning in ks.test(x = df0$ideo, y = df2$ideo): p-value will be approximate in
## the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
```

```
##
## data:  df0$ideo and df2$ideo
## D = 0.043236, p-value = 0.1917
## alternative hypothesis: two-sided
```

```
## Distribution of ideology in Add Definitions condition vs. Subtract Labels condition
ks.test(x = df1$ideo, y = df2$ideo) # reported in Appendix C
```

```
## Warning in ks.test(x = df1$ideo, y = df2$ideo): p-value will be approximate in
## the presence of ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  df1$ideo and df2$ideo
## D = 0.080759, p-value = 0.004585
## alternative hypothesis: two-sided
```

```
## Test for homogeneity of variances in Add Definitions condition vs. control
var(x = df$ideo[df$group == 0], na.rm = T) # reported in main text
```

```
## [1] 2.767426
```

```
var(x = df$ideo[df$group == 1], na.rm = T) # reported in main text
```

```
## [1] 3.218231
```

```
fligner.test(ideo ~ group,
             data = subset(df, group == 0 | group == 1)) # reported in main text
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  ideo by group
## Fligner-Killeen:med chi-squared = 13.441, df = 1, p-value = 0.0002461
```

```
## Test for homogeneity of variances in Subtract Labels condition vs. control
fligner.test(ideo ~ group,
             data = subset(df, group == 0 | group == 2)) # reported in main text
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  ideo by group
## Fligner-Killeen:med chi-squared = 0.57577, df = 1, p-value = 0.448
```

```
### Figure 4: distribution of self-reported ideology across experimental conditions ----
## General distributions
ideo_by_group <- df %>%
  filter(ideo >= 1 & ideo <= 7) %>%
  group_by(group, ideo) %>%
  summarize(count = n()) %>%
```

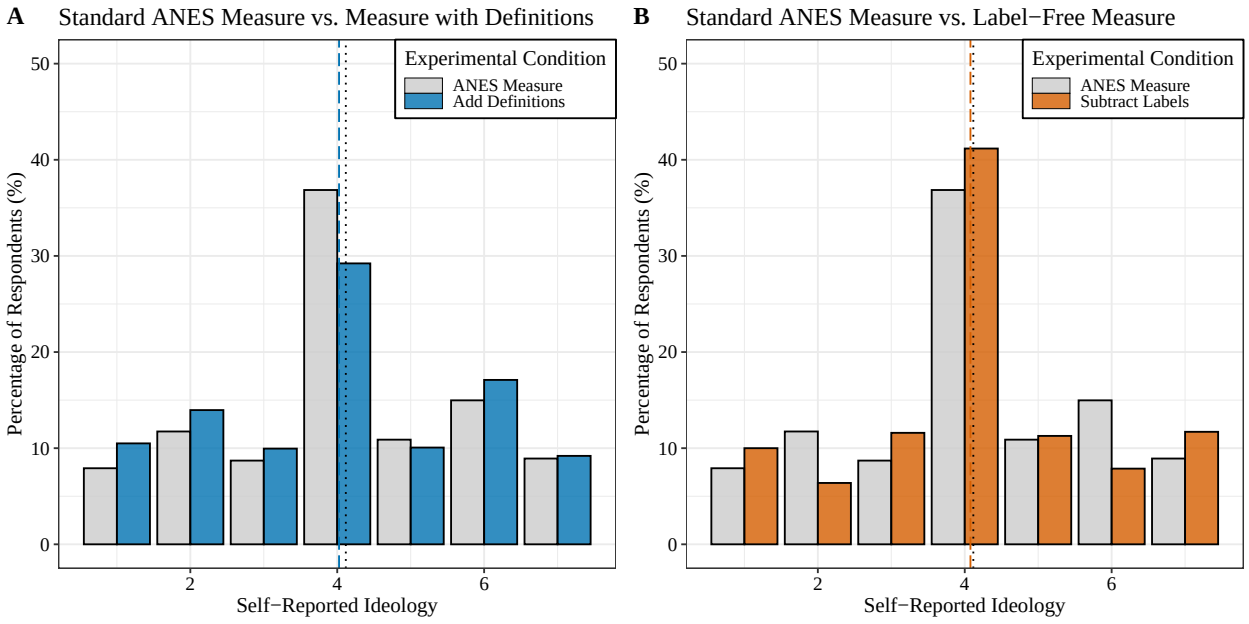
```
  mutate(freq = formattable::percent(count / sum(count)))
```

```
## `summarise()` has grouped output by 'group'. You can override using the
## `.groups` argument.
```

```
ideo_by_group$group <-
  factor(ideo_by_group$group,
         levels = c(0, 1, 2),
         labels = c("ANES Measure", "Add Definitions", "Subtract Labels"))

## Standard ANES measure vs. measure with definitions
group_mean_temp1 <- df %>%
  filter(group == 0 | group == 1) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p1 <- ggplot(data = subset(ideo_by_group, group != "Subtract Labels"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#0072B2")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#0072B2")) +
  xlab("Self-Reported Ideology") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Measure with Definitions") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p1 <- p1 +
  geom_vline(data = group_mean_temp1,
             aes(xintercept = xvalue), color = c("black", "#0072B2"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Standard ANES measure vs. label-free measure
group_mean_temp2 <- df %>%
  filter(group == 0 | group == 2) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p2 <- ggplot(data = subset(ideo_by_group, group != "Add Definitions"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
```

```
  scale_color_manual(values = c("grey80", "#D55E00")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#D55E00")) +
  xlab("Self-Reported Ideology") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Label-Free Measure") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p2 <- p2 +
  geom_vline(data = group_mean_temp2,
             aes(xintercept = xvalue), color = c("black", "#D55E00"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Combine into one graph
distribution <- plot_grid(p1, p2, labels = "AUTO", ncol = 2,
                          label_fontfamily = "Times")
```

```
# Figure 4
distribution
```




```
# ggsave(file = "distribution.pdf", width = 10, height = 5)
```

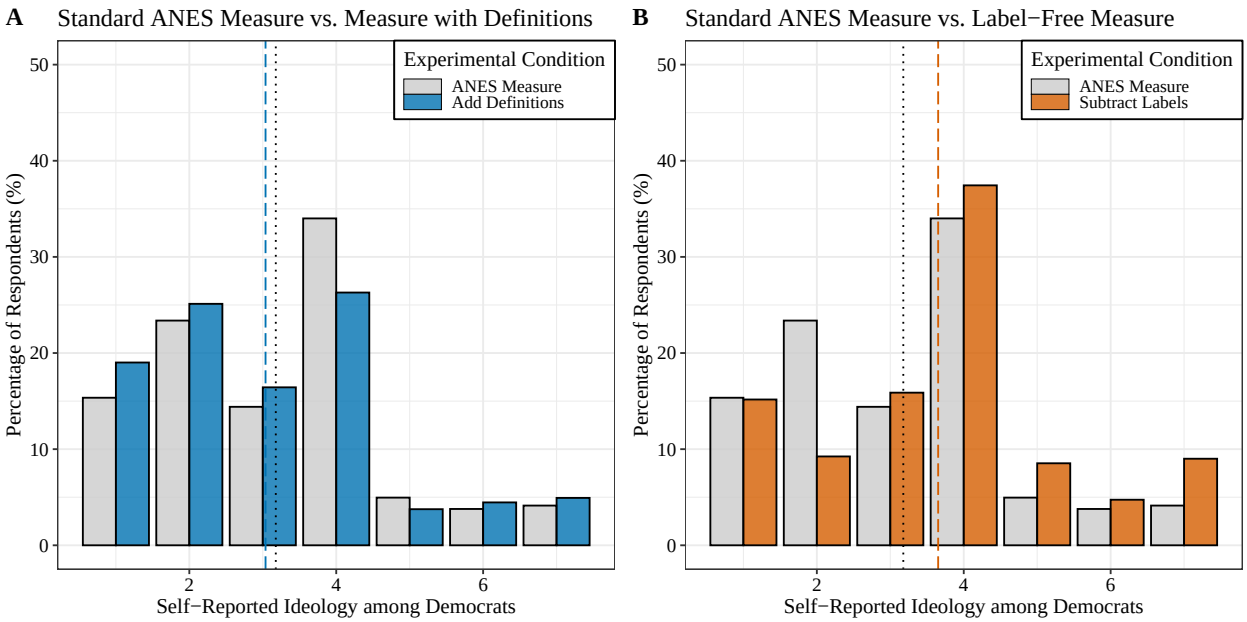
```
### Alternative analysis for Figure 4: Democrats only ----
## General distributions
ideo_by_group_dem <- df %>%
  filter(ideo >= 1 & ideo <= 7 & dem == 1) %>%
  group_by(group, ideo) %>%
  summarize(count = n()) %>%
  mutate(freq = formattable::percent(count / sum(count)))
```

```
## `summarise()` has grouped output by 'group'. You can override using the
## `.groups` argument.
```

```
ideo_by_group_dem$group <-
  factor(ideo_by_group_dem$group,
         levels = c(0, 1, 2),
         labels = c("ANES Measure", "Add Definitions", "Subtract Labels"))

## Standard ANES measure vs. measure with definitions
group_mean_temp3 <- df %>%
  filter((group == 0 | group == 1) & dem == 1) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p3 <- ggplot(data = subset(ideo_by_group_dem, group != "Subtract Labels"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#0072B2")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#0072B2")) +
  xlab("Self-Reported Ideology among Democrats") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Measure with Definitions") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p3 <- p3 +
  geom_vline(data = group_mean_temp3,
             aes(xintercept = xvalue), color = c("black", "#0072B2"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Standard ANES measure vs. label-free measure
group_mean_temp4 <- df %>%
  filter((group == 0 | group == 2) & dem == 1) %>%
```

```
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p4 <- ggplot(data = subset(ideo_by_group_dem, group != "Add Definitions"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#D55E00")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#D55E00")) +
  xlab("Self-Reported Ideology among Democrats") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Label-Free Measure") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p4 <- p4 +
  geom_vline(data = group_mean_temp4,
             aes(xintercept = xvalue), color = c("black", "#D55E00"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Combine into one graph
distribution_dem <- plot_grid(p3, p4, labels = "AUTO", ncol = 2,
                              label_fontfamily = "Times")
```

```
# Figure S8 in Appendix F
distribution_dem
```




```
# ggsave(file = "distribution_dem.pdf", width = 10, height = 5)
```

```
### Alternative analysis for Figure 4: Republicans only ----
## General distributions
ideo_by_group_gop <- df %>%
  filter(ideo >= 1 & ideo <= 7 & gop == 1) %>%
  group_by(group, ideo) %>%
  summarize(count = n()) %>%
  mutate(freq = formattable::percent(count / sum(count)))
```

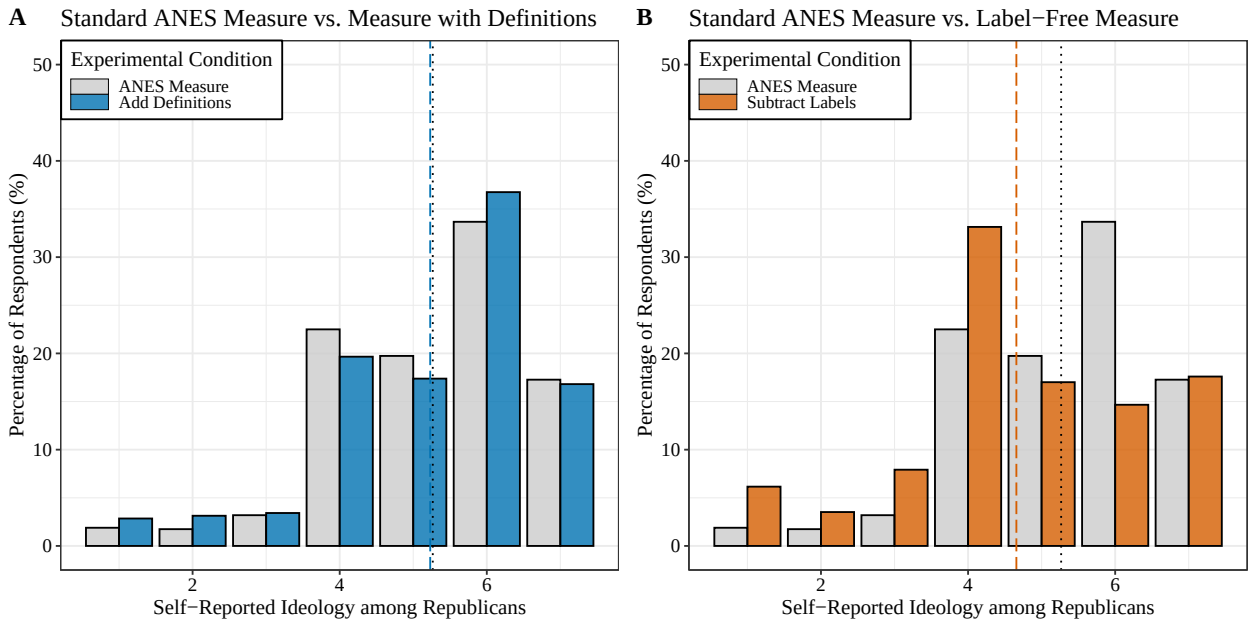
```
## `summarise()` has grouped output by 'group'. You can override using the
## `.groups` argument.
```

```
ideo_by_group_gop$group <-
  factor(ideo_by_group_gop$group,
         levels = c(0, 1, 2),
         labels = c("ANES Measure", "Add Definitions", "Subtract Labels"))

## Standard ANES measure vs. measure with definitions
group_mean_temp5 <- df %>%
  filter((group == 0 | group == 1) & gop == 1) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p5 <- ggplot(data = subset(ideo_by_group_gop, group != "Subtract Labels"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#0072B2")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#0072B2")) +
  xlab("Self-Reported Ideology among Republicans") +
  ylab("Percentage of Respondents (%)") +
```

```
  ggtitle("Standard ANES Measure vs. Measure with Definitions") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(0.397, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p5 <- p5 +
  geom_vline(data = group_mean_temp5,
             aes(xintercept = xvalue), color = c("black", "#0072B2"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Standard ANES measure vs. label-free measure
group_mean_temp6 <- df %>%
  filter((group == 0 | group == 2) & gop == 1) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p6 <- ggplot(data = subset(ideo_by_group_gop, group != "Add Definitions"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#D55E00")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#D55E00")) +
  xlab("Self-Reported Ideology among Republicans") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Label-Free Measure") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(0.397, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p6 <- p6 +
  geom_vline(data = group_mean_temp6,
             aes(xintercept = xvalue), color = c("black", "#D55E00"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Combine into one graph
distribution_gop <- plot_grid(p5, p6, labels = "AUTO", ncol = 2,
```

```
                                    label_fontfamily = "Times")
```

```
# Figure S9 in Appendix F
distribution_gop
```




```
# ggsave(file = "distribution_gop.pdf", width = 10, height = 5)
```

```
### Alternative analysis for Figure 4: Independents only ----
## General distributions
ideo_by_group_ind <- df %>%
  filter(ideo >= 1 & ideo <= 7 & (dem == 0 & gop == 0)) %>%
  group_by(group, ideo) %>%
  summarize(count = n()) %>%
  mutate(freq = formattable::percent(count / sum(count)))
```

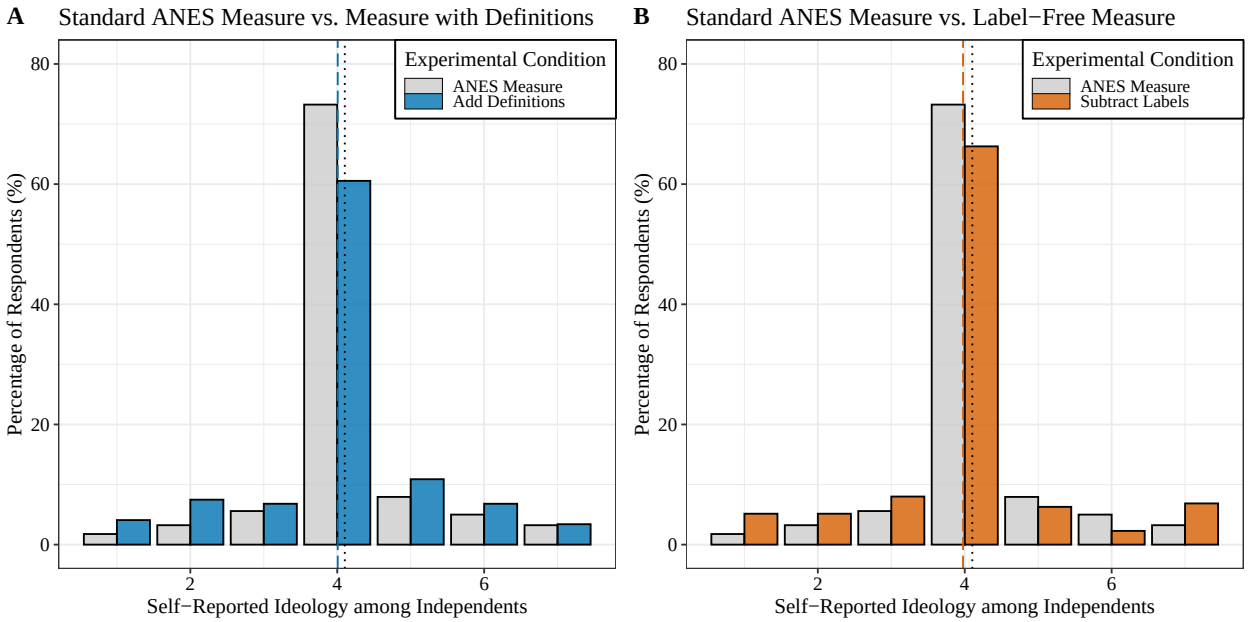
```
## `summarise()` has grouped output by 'group'. You can override using the
## `.groups` argument.
```

```
ideo_by_group_ind$group <-
  factor(ideo_by_group_ind$group,
         levels = c(0, 1, 2),
         labels = c("ANES Measure", "Add Definitions", "Subtract Labels"))

## Standard ANES measure vs. measure with definitions
group_mean_temp7 <- df %>%
  filter((group == 0 | group == 1) & (dem == 0 & gop == 0)) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p7 <- ggplot(data = subset(ideo_by_group_ind, group != "Subtract Labels"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
```

```
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#0072B2")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#0072B2")) +
  xlab("Self-Reported Ideology among Independents") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Measure with Definitions") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 80))
p7 <- p7 +
  geom_vline(data = group_mean_temp7,
             aes(xintercept = xvalue), color = c("black", "#0072B2"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Standard ANES measure vs. label-free measure
group_mean_temp8 <- df %>%
  filter((group == 0 | group == 2) & (dem == 0 & gop == 0)) %>%
  group_by(group) %>%
  summarize(xvalue = mean(ideo, na.rm = T))
p8 <- ggplot(data = subset(ideo_by_group_ind, group != "Add Definitions"),
             aes(x = ideo, y = freq * 100, color = group, fill = group)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.8) +
  scale_color_manual(values = c("grey80", "#D55E00")) +
  scale_fill_manual("Experimental Condition", values = c("grey80", "#D55E00")) +
  xlab("Self-Reported Ideology among Independents") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Standard ANES Measure vs. Label-Free Measure") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 80))
p8 <- p8 +
```

```r
  geom_vline(data = group_mean_temp8,
             aes(xintercept = xvalue), color = c("black", "#D55E00"),
             linetype = c("dotted", "longdash"), size = 0.5, show.legend = F)

## Combine into one graph
distribution_ind <- plot_grid(p7, p8, labels = "AUTO", ncol = 2,
                              label_fontfamily = "Times")
```

```r
# Figure S10 in Appendix F
distribution_ind
```




```r
# ggsave(file = "distribution_ind.pdf", width = 10, height = 5)
```

```r
### Table 2: treatment effects on self-reported ideology by subgroup ----
## Regression analysis
mod0 <- lm_robust(ideo ~ group,
                  data = df)
mod1 <- lm_robust(ideo ~ group + gop + dem + black + female + sophis,
                  data = df)
mod2 <- lm_robust(ideo ~ group + gop + dem + black + female + sophis +
                    group * (gop + dem),
                  data = df)
mod3 <- lm_robust(ideo ~ group + gop + dem + black + female + sophis +
                    group * black,
                  data = df)
mod4 <- lm_robust(ideo ~ group + gop + dem + black + female + sophis +
                    group * female,
                  data = df)
mod5 <- lm_robust(ideo ~ group + gop + dem + black + female + sophis +
```

```
                    group * sophis,
                  data = df)
mod6 <- lm_robust(ideo ~ group * (gop + dem + black + female + sophis),
                  data = df)
texreg(list(mod0, mod1, mod2, mod3, mod4, mod5, mod6),
       stars = 0,
       include.ci = F,
       custom.header = list("Dependent Variable: Self-Reported Ideology" = 1:7),
       custom.coef.names =
         c("Constant", "Add Definitions", "Subtract Labels",
           "Republican", "Democrat", "Black", "Female", "Politically Sophisticated (PS)",
           "Add Definitions × Republican", "Subtract Labels × Republican",
           "Add Definitions × Democrat", "Subtract Labels × Democrat",
           "Add Definitions × Black", "Subtract Labels × Black",
           "Add Definitions × Female", "Subtract Labels × Female",
           "Add Definitions × PS", "Subtract Labels × PS"),
       custom.note = "Entries are OLS estimates with robust standard errors in parentheses.",
       caption = "Self-Reported Ideologies in Different Social Groups Are Changed by Question
       fontsize = "small"
)
```

```
##
## \begin{table}
## \begin{center}
## \begin{small}
## \begin{tabular}{l c c c c c c c}
## \hline
##  & \multicolumn{7}{c}{Dependent Variable: Self-Reported Ideology} \\
## \cline{2-8}
##  & Model 1 & Model 2 & Model 3 & Model 4 & Model 5 & Model 6 & Model 7 \\
## \hline
## Constant                       & $4.12$   & $4.16$   & $4.18$   & $4.17$   & $4.11$   & $4.1
##                                & $(0.04)$ & $(0.06)$ & $(0.07)$ & $(0.06)$ & $(0.07)$ & $(0
## Add Definitions                & $-0.09$  & $-0.10$  & $-0.11$  & $-0.11$  & $0.03$   & $-0
##                                & $(0.07)$ & $(0.06)$ & $(0.11)$ & $(0.06)$ & $(0.09)$ & $(0
## Subtract Labels                & $-0.04$  & $-0.02$  & $-0.12$  & $-0.06$  & $0.02$   & $0.
##                                & $(0.07)$ & $(0.06)$ & $(0.11)$ & $(0.07)$ & $(0.10)$ & $(0
## Republican                     &          & $1.04$   & $1.13$   & $1.04$   & $1.04$   & $1.0
##                                &          & $(0.06)$ & $(0.07)$ & $(0.06)$ & $(0.06)$ & $(0
## Democrat                       &          & $-0.80$  & $-0.94$  & $-0.80$  & $-0.80$  & $-0
##                                &          & $(0.06)$ & $(0.07)$ & $(0.06)$ & $(0.06)$ & $(0
## Black                          &          & $0.05$   & $0.05$   & $-0.04$  & $0.05$   & $0.0
##                                &          & $(0.08)$ & $(0.08)$ & $(0.10)$ & $(0.08)$ & $(0
## Female                         &          & $-0.18$  & $-0.17$  & $-0.18$  & $-0.10$  & $-0
##                                &          & $(0.05)$ & $(0.05)$ & $(0.05)$ & $(0.07)$ & $(0
## Politically Sophisticated (PS) &          & $0.06$   & $0.06$   & $0.06$   & $0.06$   & $0.
##                                &          & $(0.05)$ & $(0.05)$ & $(0.05)$ & $(0.05)$ & $(0
```

```
## Add Definitions × Republican   &          &          & $0.08$   &          &          &
##                                &          &          & $(0.15)$ &          &          &
## Subtract Labels × Republican   &          &          & $-0.45$  &          &          &
##                                &          &          & $(0.15)$ &          &          &
## Add Definitions × Democrat     &          &          & $-0.04$  &          &          &
##                                &          &          & $(0.15)$ &          &          &
## Subtract Labels × Democrat     &          &          & $0.58$   &          &          &
##                                &          &          & $(0.15)$ &          &          &
## Add Definitions × Black        &          &          &          & $0.05$   &          &
##                                &          &          &          & $(0.19)$ &          &
## Subtract Labels × Black        &          &          &          & $0.31$   &          &
##                                &          &          &          & $(0.19)$ &          &
## Add Definitions × Female       &          &          &          &          & $-0.25$  &
##                                &          &          &          &          & $(0.12)$ &
## Subtract Labels × Female       &          &          &          &          & $-0.06$  &
##                                &          &          &          &          & $(0.13)$ &
## Add Definitions × PS           &          &          &          &          &          & $0.0
##                                &          &          &          &          &          & $(0
## Subtract Labels × PS           &          &          &          &          &          & $-0
##                                &          &          &          &          &          & $(0
## \hline
## R$^2$                          & $0.00$   & $0.25$   & $0.26$   & $0.25$   & $0.25$   & $0.2
## Adj. R$^2$                     & $-0.00$  & $0.25$   & $0.26$   & $0.25$   & $0.25$   & $0.2
## Num. obs.                      & $3747$   & $3669$   & $3669$   & $3669$   & $3669$   & $360
## RMSE                           & $1.70$   & $1.47$   & $1.46$   & $1.47$   & $1.47$   & $1.4
## \hline
## \multicolumn{8}{l}{\tiny{Entries are OLS estimates with robust standard errors in parenthese
## \end{tabular}
## \end{small}
## \caption{Self-Reported Ideologies in Different Social Groups Are Changed by Question Wording
## \label{table:coefficients}
## \end{center}
## \end{table}
```

```
### Figure 5: within-subjects differences in ideology ----
## Construct the variable that indicates within-subjects differences in ideology
df$ideo_change <- df$ideo - df$ideo2

## Formally test the within-subject differences by partisanship
# Democrats (reported in main text)
lm_robust(ideo_change ~ 1, data = df, subset = (group == 2 & dem == 1))
```

```
##               Estimate Std. Error  t value     Pr(>|t|)  CI Lower CI Upper  DF
## (Intercept) 0.5909091  0.1004004 5.885527 8.162034e-09 0.3935552 0.788263 417
```

```
# Independents (reported in Footnote 19)
lm_robust(ideo_change ~ 1, data = df, subset = (group == 2 & dem == 0 & gop == 0))
```

```
##               Estimate Std. Error   t value  Pr(>|t|)   CI Lower  CI Upper  DF
```

```
## (Intercept) 0.03488372  0.1058905 0.3294321 0.7422323 -0.1741371 0.2439045 171
```

```
# Republicans (reported in main text)
lm_robust(ideo_change ~ 1, data = df, subset = (group == 2 & gop == 1))
```

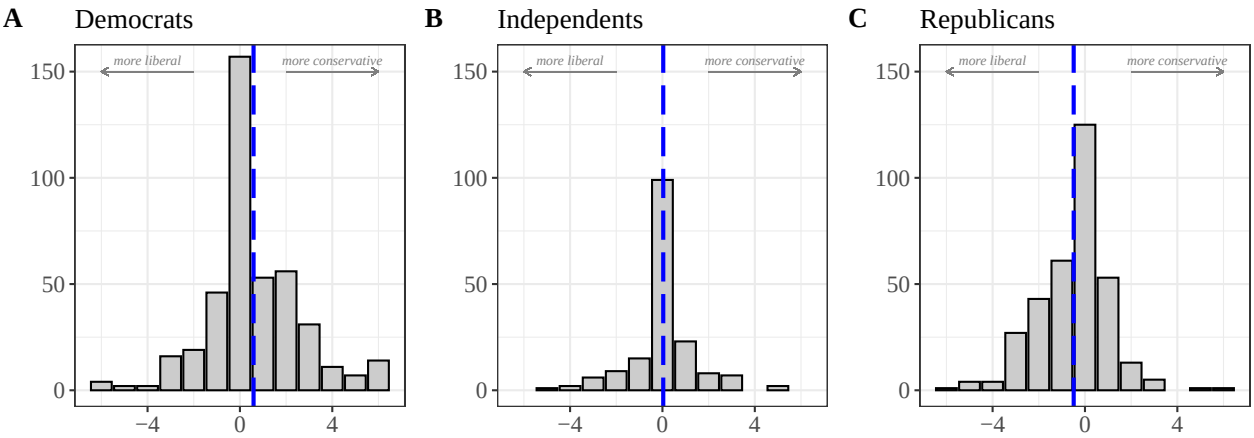
```
##                Estimate Std. Error   t value     Pr(>|t|)   CI Lower   CI Upper
## (Intercept) -0.4881657 0.08414264 -5.801644 1.514089e-08 -0.6536766 -0.3226547
##                DF
## (Intercept) 337
```

```
## Distribution of within-subjects differences in ideology by partisanship
# Democrats
p1 <- ggplot(subset(df, group == 2 & dem == 1), aes(x = ideo_change)) +
  geom_bar(color = "black", fill = "grey80", na.rm = T) +
  theme_bw() +
  ggtitle("Democrats") +
  xlab("") +
  ylab("") +
  coord_cartesian(xlim = c(-6.5, 6.5), ylim = c(0, 155)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

# Independents
p2 <- ggplot(subset(df, group == 2 & dem == 0 & gop == 0), aes(x = ideo_change)) +
  geom_bar(color = "black", fill = "grey80", na.rm = T) +
  theme_bw() +
  ggtitle("Independents") +
  xlab("") +
  ylab("") +
  coord_cartesian(xlim = c(-6.5, 6.5), ylim = c(0, 155)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

# Republicans
p3 <- ggplot(subset(df, group == 2 & gop == 1), aes(x = ideo_change)) +
  geom_bar(color = "black", fill = "grey80", na.rm = T) +
  theme_bw() +
  ggtitle("Republicans") +
  xlab("") +
  ylab("") +
  coord_cartesian(xlim = c(-6.5, 6.5), ylim = c(0, 155)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 +
```

```
  geom_vline(xintercept = mean(subset(df, group == 2 & dem == 1)$ideo_change, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)
p2 <- p2 +
  geom_vline(xintercept = mean(subset(df, group == 2 & dem == 0 & gop == 0)$ideo_change, na.rm
             linetype = "longdash", color = "blue", size = 1)
p3 <- p3 +
  geom_vline(xintercept = mean(subset(df, group == 2 & gop == 1)$ideo_change, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)

## Add annotations (left arrow = more liberal; right arrow = more conservative)
p1 <- p1 +
  geom_segment(aes(x = -2, y = 150, xend = -6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = -4, y = 155, label = "more liberal",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times") +
  geom_segment(aes(x = 2, y = 150, xend = 6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = 4, y = 155, label = "more conservative",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times")
p2 <- p2 +
  geom_segment(aes(x = -2, y = 150, xend = -6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = -4, y = 155, label = "more liberal",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times") +
  geom_segment(aes(x = 2, y = 150, xend = 6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = 4, y = 155, label = "more conservative",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times")
p3 <- p3 +
  geom_segment(aes(x = -2, y = 150, xend = -6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = -4, y = 155, label = "more liberal",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times") +
  geom_segment(aes(x = 2, y = 150, xend = 6, yend = 150),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = 4, y = 155, label = "more conservative",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times")

## Combine into one graph
ideo_change_pid <- plot_grid(p1, p2, p3, labels = "AUTO", ncol = 3,
                             label_fontfamily = "Times")
```

```
ideo_change_pid <-
  ggdraw(add_sub(ideo_change_pid,
                 "Within-Subjects Difference in Ideology (Label-Free Measure - ANES Measure)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```
# Figure 5
ideo_change_pid
```




```
# ggsave(file = "ideo_change_pid.pdf", width = 9, height = 3.5)
```

```
### Figure 6: magnitude of within-subject effects vis-a-vis baseline ideological
### differences between partisans ----
## Calculate the baseline ideological differences between Democrats and Republicans
lm_robust(ideo2 ~ dem, data = df,
          subset = group == 2 & (dem == 1 | gop == 1) & (ideo >= 1 & ideo <= 7))
```

```
##              Estimate Std. Error   t value     Pr(>|t|)  CI Lower  CI Upper  DF
## (Intercept)  5.174556  0.0741749  69.76155 0.000000e+00  5.028942  5.320170 754
## dem         -2.117140  0.1055427 -20.05956 4.682563e-72 -2.324332 -1.909948 754
```

```
# diff. = 2.12, SE = 0.11

## Calculate the ideological differences between Democrats and Republicans when
## ideological labels are removed
lm_robust(ideo ~ dem, data = df,
          subset = group == 2 & (dem == 1 | gop == 1) & (ideo2 >= 1 & ideo2 <= 7))
```

```
##              Estimate Std. Error   t value      Pr(>|t|)  CI Lower   CI Upper
## (Intercept)  4.686391  0.0879336 53.294652 6.539945e-258  4.513767  4.8590143
## dem         -1.038065  0.1206933 -8.600852  4.554909e-17 -1.275000 -0.8011303
##              DF
## (Intercept) 754
## dem         754
```

```r
# diff. = 1.04, SE = 0.12

## Indicate respondents' partisanship
df2 <- df2 %>% mutate(party = case_when(
  pid1 == 1 & pid2d == 1 ~ "Democrats",
  pid1 == 1 & pid2d == 2 ~ "Democrats",
  (pid1 == 3 | pid1 == 4) & pid2n == 2 ~ "Democrats",
  (pid1 == 3 | pid1 == 4) & pid2n == 1 ~ "Republicans",
  pid1 == 2 & pid2r == 2 ~ "Republicans",
  pid1 == 2 & pid2r == 1 ~ "Republicans"))

## Obtain the distribution of self-reported ideology among partisans in the ANES measure
ideo_by_party_ANES <- df2 %>%
  filter(ideo >= 1 & ideo <= 7) %>%
  filter(ideo2 >= 1 & ideo2 <= 7) %>%
  filter(party == "Democrats" | party == "Republicans") %>%
  group_by(party, ideo2) %>%
  summarize(count_ANES = n()) %>%
  mutate(freq_ANES = formattable::percent(count_ANES / sum(count_ANES))) %>%
  rename(ideo = ideo2)
```

```
## `summarise()` has grouped output by 'party'. You can override using the
## `.groups` argument.
```

```r
## Obtain the distribution of self-reported ideology among partisans in the label-free measure
ideo_by_party_nolabel <- df2 %>%
  filter(ideo >= 1 & ideo <= 7) %>%
  filter(ideo2 >= 1 & ideo2 <= 7) %>%
  filter(party == "Democrats" | party == "Republicans") %>%
  group_by(party, ideo) %>%
  summarize(count_nolabel = n()) %>%
  mutate(freq_nolabel = formattable::percent(count_nolabel / sum(count_nolabel)))
```

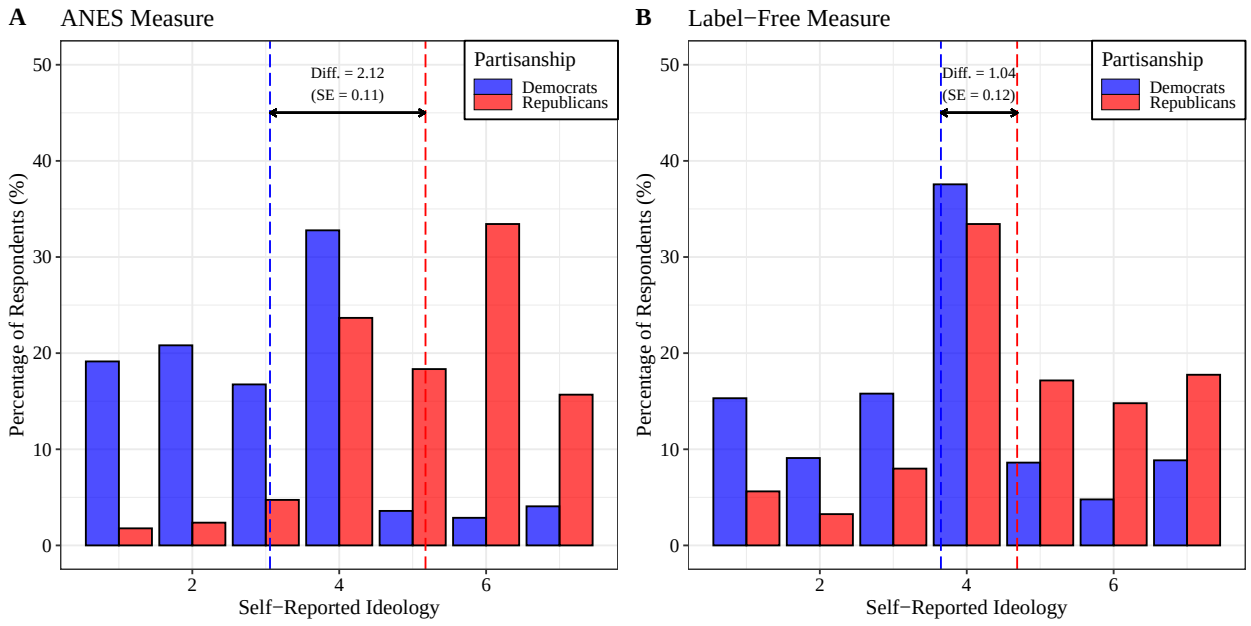
```
## `summarise()` has grouped output by 'party'. You can override using the
## `.groups` argument.
```

```r
## Merge the results
ideo_by_party <- merge(ideo_by_party_ANES, ideo_by_party_nolabel, by = c("ideo", "party"))

## Standard ANES measure in Subtract Labels condition
group_mean_temp1 <- df2 %>%
  filter(ideo >= 1 & ideo <= 7) %>%
  filter(ideo2 >= 1 & ideo2 <= 7) %>%
  filter(party == "Democrats" | party == "Republicans") %>%
  group_by(party) %>%
  summarize(xvalue = mean(ideo2))
p1 <- ggplot(data = ideo_by_party, aes(x = ideo, y = freq_ANES * 100,
                                       color = party, fill = party)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
```

```
          color = "black", alpha = 0.7) +
  scale_color_manual(values = c("blue", "red")) +
  scale_fill_manual("Partisanship", values = c("blue", "red")) +
  xlab("Self-Reported Ideology") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("ANES Measure") +
  theme_bw() +
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p1 <- p1 +
  geom_vline(data = group_mean_temp1,
             aes(xintercept = xvalue), color = c("blue", "red"),
             linetype = c("longdash", "longdash"), size = 0.5, show.legend = F)
p1 <- p1 +
  geom_segment(aes(x = 4.116, y = 45, xend = 3.06, yend = 45),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .5, color = "black") +
  geom_segment(aes(x = 4.116, y = 45, xend = 5.17, yend = 45),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .5, color = "black") +
  annotate("text", x = 4.116, y = 48, label = "Diff. = 2.12\n(SE = 0.11)",
           size = 3, color = "black", family = "Times")

## Label-free measure in Subtract Labels condition
group_mean_temp2 <- df2 %>%
  filter(ideo >= 1 & ideo <= 7) %>%
  filter(ideo2 >= 1 & ideo2 <= 7) %>%
  filter(party == "Democrats" | party == "Republicans") %>%
  group_by(party) %>%
  summarize(xvalue = mean(ideo))
p2 <- ggplot(data = ideo_by_party, aes(x = ideo, y = freq_nolabel * 100,
                                       color = party, fill = party)) +
  geom_bar(stat = "identity", position = position_dodge(.9),
           color = "black", alpha = 0.7) +
  scale_color_manual(values = c("blue", "red")) +
  scale_fill_manual("Partisanship", values = c("blue", "red")) +
  xlab("Self-Reported Ideology") +
  ylab("Percentage of Respondents (%)") +
  ggtitle("Label-Free Measure") +
  theme_bw() +
```

```
  theme(text = element_text(color = "black", size = 12, family = "Times"),
        axis.text = element_text(color = "black", family = "Times", size = 11),
        legend.justification = c(1, 1),
        legend.position = c(1, 1),
        legend.background = element_rect(fill = "white", color = "black"),
        legend.box.background = element_rect(color = "black"),
        legend.key.size = unit(1.5, "line"),
        legend.key.height = unit(0, "cm")) +
  coord_cartesian(ylim = c(0, 50))
p2 <- p2 +
  geom_vline(data = group_mean_temp2,
             aes(xintercept = xvalue), color = c("blue", "red"),
             linetype = c("longdash", "longdash"), size = 0.5, show.legend = F)
p2 <- p2 +
  geom_segment(aes(x = 4.167, y = 45, xend = 3.65, yend = 45),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .5, color = "black") +
  geom_segment(aes(x = 4.167, y = 45, xend = 4.69, yend = 45),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .5, color = "black") +
  annotate("text", x = 4.167, y = 48, label = "Diff. = 1.04\n(SE = 0.12)",
           size = 3, color = "black", family = "Times")

## Combine into one graph
ideo_close_gap <- plot_grid(p1, p2, labels = "AUTO", ncol = 2,
                            label_fontfamily = "Times")
```

```
# Figure 6
ideo_close_gap
```



```r
# ggsave(file = "ideo_close_gap.pdf", width = 10, height = 5)
```

```r
### Analysis: proportion of respondents in Dem/Rep Party moving from lib/con to con/lib ----
## Proportion of Democrats moving from liberal to conservative
df2$moved_dem <- ifelse((df2$ideo2 >= 1 & df2$ideo2 <= 3) &
                          (df2$ideo >= 5 & df2$ideo <= 7) & df2$dem == 1, 1, 0)
table(df2$moved_dem) # 6% Democrats shifted, as reported in main text
```

```
##
##   0   1
## 886  52
```

```r
## Proportion of Republicans moving from conservative to liberal
df2$moved_gop <- ifelse((df2$ideo2 >= 5 & df2$ideo2 <= 7) &
                          (df2$ideo >= 1 & df2$ideo <= 3) & df2$gop == 1, 1, 0)
table(df2$moved_gop) # 3% Republicans shifted, as reported in main text
```

```
##
##   0   1
## 910  26
```

```r
### Figure 7: implications for partisan-ideological sorting in the US ----
## Test for equality of correlation coefficients (control vs. Add Definitions)
cor.test.plus(cor.test(df0$pid, df0$ideo))
```

```
## [[1]]
##
##  Pearson's product-moment correlation
##
## data:  df0$pid and df0$ideo
## t = 30.918, df = 1874, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5504238 0.6104035
## sample estimates:
##       cor
## 0.5812025
##
##
## $Standard.Error
## [1] 0.01879797
```

```r
cor.test.plus(cor.test(df1$pid, df1$ideo))
```

```
## [[1]]
##
##  Pearson's product-moment correlation
##
## data:  df1$pid and df1$ideo
## t = 20.655, df = 922, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5166841 0.6049841
## sample estimates:
##       cor
## 0.5624356
##
##
## $Standard.Error
## [1] 0.02723056
```

```
res <- twocorci(df0$pid, df0$ideo, df1$pid, df1$ideo, method = "pearson")
res$difference
```

```
## [1] 0.01876693
```

```
res$p.value
```

```
## [1] 0.584
```

```
## Test for equality of correlation coefficients (control vs. Subtract Labels)
cor.test.plus(cor.test(df2$pid, df2$ideo))
```

```
## [[1]]
##
##  Pearson's product-moment correlation
##
## data:  df2$pid and df2$ideo
## t = 8.2131, df = 936, p-value = 7.114e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1985603 0.3180073
## sample estimates:
##      cor
## 0.259275
##
##
## $Standard.Error
## [1] 0.03156828
```

```
res <- twocorci(df0$pid, df0$ideo, df2$pid, df2$ideo, method = "pearson")
res$difference
```

```
## [1] 0.3219276
```

```
res$p.value
```

```
## [1] 0
```

```
## Test for equality of correlation coefficients (within-group for Subtract Labels)
cor.test.plus(cor.test(df2$pid, df2$ideo2))
```

```
## [[1]]
##
##  Pearson's product-moment correlation
##
## data:  df2$pid and df2$ideo2
## t = 21.705, df = 926, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5363894 0.6218152
## sample estimates:
##       cor
## 0.5806985
##
##
## $Standard.Error
## [1] 0.0267536
```

```
res <- twocorci(df2$pid, df2$ideo2, df2$pid, df2$ideo, method = "pearson")
res$estimate1
```

```
## [1] 0.5806985
```

```
res$estimate2
```

```
## [1] 0.259275
```

```
res$difference
```

```
## [1] 0.3214235
```

```
res$p.value
```

```
## [1] 0
```

```
## Test for equality of correlation coefficients (Add Definitions vs. Subtract Labels)
res <- twocorci(df1$pid, df1$ideo, df2$pid, df2$ideo, method = "pearson")
res$estimate1
```

```
## [1] 0.5624356
```

```
res$estimate2
```

```
## [1] 0.259275
```

```
res$difference
```

```
## [1] 0.3031606
```

```
res$p.value
```

```
## [1] 0
```

```
## Count the number of observations for each group (as reported in main text)
df0 %>%
  dplyr::select(ideo, pid) %>%
```

```
  filter(rowSums(is.na(.)) == 0) %>%
  count()
```

```
##      n
## 1 1876
```

```
df1 %>%
  dplyr::select(ideo, pid) %>%
  filter(rowSums(is.na(.)) == 0) %>%
  count()
```

```
##     n
## 1 924
```

```
df2 %>%
  dplyr::select(ideo, pid) %>%
  filter(rowSums(is.na(.)) == 0) %>%
  count()
```

```
##     n
## 1 938
```

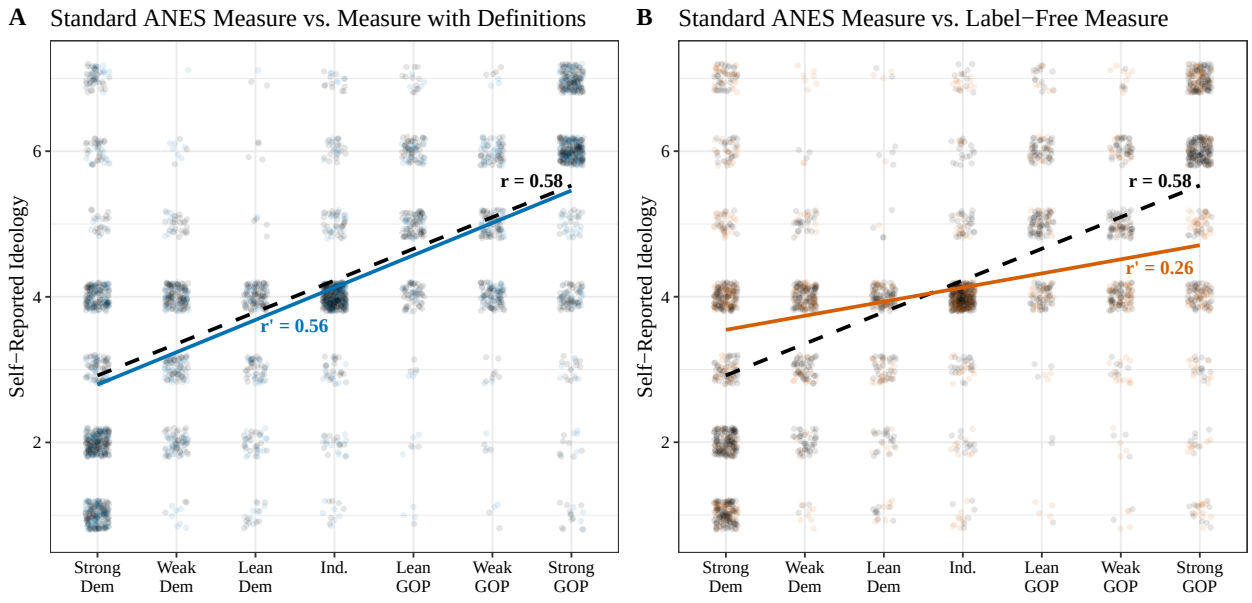
```
df2 %>%
  dplyr::select(ideo2, pid) %>%
  filter(rowSums(is.na(.)) == 0) %>%
  count()
```

```
##     n
## 1 928
```

```
## Correlation between partisanship and ideology by experimental condition
# Standard ANES measure vs. measure with definitions
p1 <- ggplot(subset(df, group == 0 | group == 1),
             aes(x = pid, y = ideo, color = group, linetype = group)) +
  geom_jitter(position = position_jitter(width = .15, height = .2, seed = 1234567),
              size = 1, alpha = .1, na.rm = T) +
  stat_smooth(method = "lm", formula = y ~ x, se = F, na.rm = T) +
  scale_color_manual(values = c("black", "#0072B2")) +
  scale_linetype_manual(values = c("dashed", "solid")) +
  scale_x_discrete(
    "",
    limits = c("Strong\nDem", "Weak\nDem", "Lean\nDem", "Ind.",
               "Lean\nGOP", "Weak\nGOP", "Strong\nGOP")
  ) +
  theme_bw() +
  ggtitle("Standard ANES Measure vs. Measure with Definitions") +
  ylab("Self-Reported Ideology") +
  annotate("text", x = 6.5, y = 5.60, label = "r = 0.58",
           size = 4, color = "black", fontface = "bold", family = "Times") +
  annotate("text", x = 3.5, y = 3.6, label = "r' = 0.56",
```

```
           size = 4, color = "#0072B2", fontface = "bold", family = "Times") +
  theme(text = element_text(color = "black", family = "Times", size = 12),
        axis.text = element_text(color = "black", family = "Times", size = 10),
        legend.position = "none")

# Standard ANES measure vs. label-free measure
p2 <- ggplot(subset(df, group == 0 | group == 2),
             aes(x = pid, y = ideo, color = group, linetype = group)) +
  geom_jitter(position = position_jitter(width = .15, height = .2, seed = 1234567),
              size = 1, alpha = .1, na.rm = T) +
  stat_smooth(method = "lm", formula = y ~ x, se = F, na.rm = T) +
  scale_color_manual(values = c("black", "#D55E00")) +
  scale_linetype_manual(values = c("dashed", "solid")) +
  scale_x_discrete(
    "",
    limits = c("Strong\nDem", "Weak\nDem", "Lean\nDem", "Ind.",
               "Lean\nGOP", "Weak\nGOP", "Strong\nGOP")
  ) +
  theme_bw() +
  ggtitle("Standard ANES Measure vs. Label-Free Measure") +
  ylab("Self-Reported Ideology") +
  annotate("text", x = 6.5, y = 5.60, label = "r = 0.58",
           size = 4, color = "black", fontface = "bold", family = "Times") +
  annotate("text", x = 6.5, y = 4.40, label = "r' = 0.26",
           size = 4, color = "#D55E00", fontface = "bold", family = "Times") +
  theme(text = element_text(color = "black", family = "Times", size = 12),
        axis.text = element_text(color = "black", family = "Times", size = 10),
        legend.position = "none")

## Combine into one graph
sorting <- plot_grid(p1, p2, labels = "AUTO", ncol = 2,
                     label_fontfamily = "Times")
```

```
# Figure 7
sorting
```




```
# ggsave(file = "sorting.pdf", width = 10, height = 5)
```

```
### Analysis: correlation between fiscal preferences and ideology ----
## Correlation between fiscal preferences and ideology
# For the control group (as reported in Appendix C)
corci(df0$fiscal, df0$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.3152463 0.4131219
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.3659849
```

```
# For Add Definitions condition (as reported in Appendix C)
corci(df1$fiscal, df1$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.3499404 0.4734404
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.4145346
```

```
# For Subtract Labels condition (as reported in Appendix C)
corci(df2$fiscal, df2$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.2276916 0.3853732
```

```
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.3079472
```

```
# Control group vs. Add Definitions condition (as reported in Appendix C)
res <- twocorci(df0$fiscal, df0$ideo, df1$fiscal, df1$ideo, method = "pearson")
res$difference
```

```
## [1] -0.04854968
```

```
res$p.value
```

```
## [1] 0.204
```

```
# Control group vs. Subtract Labels condition (as reported in Appendix C)
res <- twocorci(df0$fiscal, df0$ideo, df2$fiscal, df2$ideo, method = "pearson")
res$difference
```

```
## [1] 0.05803766
```

```
res$p.value
```

```
## [1] 0.216
```

```
# Add Definitions condition vs. Subtract Labels condition (as reported in Appendix C)
res <- twocorci(df1$fiscal, df1$ideo, df2$fiscal, df2$ideo, method = "pearson")
res$difference
```

```
## [1] 0.1065873
```

```
res$p.value
```

```
## [1] 0.037
```

```
### Analysis: correlation between government responsibility and ideology ----
## Correlation between government responsibility and ideology
# For the control group (as reported in Appendix C)
corci(df0$respon, df0$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.2531228 0.3498100
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.2982743
```

```
# For Add Definitions condition (as reported in Appendix C)
corci(df1$respon, df1$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.3207242 0.4514131
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.3874579
```

```
# For Subtract Labels condition (as reported in Appendix C)
corci(df2$respon, df2$ideo, method = "pearson", saveboot = F)
```

```
## $conf.int
## [1] 0.4282931 0.5505039
##
## $p.value
## [1] 0
##
## $estimate
## [1] 0.4924842
```

```
# Control group vs. Add Definitions condition (as reported in Appendix C)
res <- twocorci(df0$respon, df0$ideo, df1$respon, df1$ideo, method = "pearson")
res$difference
```

```
## [1] -0.08918361
```

```
res$p.value
```

```
## [1] 0.036
```

```
# Control group vs. Subtract Labels condition (as reported in Appendix C)
res <- twocorci(df0$respon, df0$ideo, df2$respon, df2$ideo, method = "pearson")
res$difference
```

```
## [1] -0.1942098
```

```
res$p.value
```

```
## [1] 0
```

```
# Add Definitions condition vs. Subtract Labels condition (as reported in Appendix C)
res <- twocorci(df1$respon, df1$ideo, df2$respon, df2$ideo, method = "pearson")
res$difference
```

```
## [1] -0.1050262
```

```
res$p.value
```

```
## [1] 0.02
```

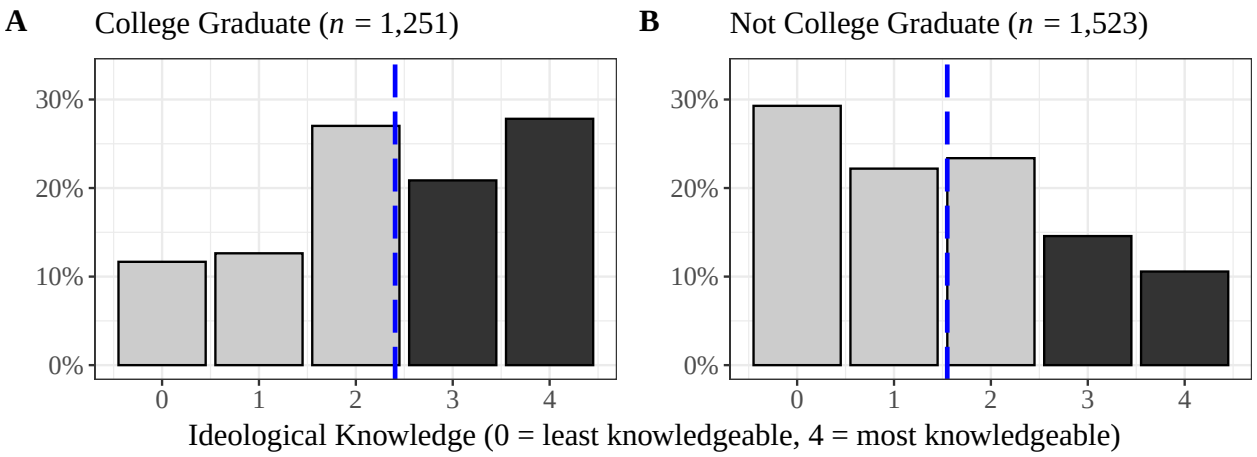
```
### Analysis: heterogeneity by education ----
## Distribution of ideological knowledge by education
table(df$college, df$ideo_know)
```

```
##
##       0   1   2   3   4
##   0 446 338 356 222 161
##   1 146 158 338 261 348
```

```
p1 <- ggplot(subset(df, college == 1), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("College Graduate (", italic("n"), " = 1,251)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .33)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")
p2 <- ggplot(subset(df, college == 0), aes(x = ideo_know, fill = ideo_know >= 3)) +
  geom_bar(aes(y = after_stat(count) / sum(after_stat(count))),
           color = "black", na.rm = T) +
  scale_fill_manual(values = c("grey80", "grey20")) +
  scale_y_continuous(labels = scales::percent) +
  theme_bw() +
  ggtitle(expression(paste("Not College Graduate (", italic("n"), " = 1,523)"))) +
  xlab("") +
  ylab("") +
  coord_cartesian(ylim = c(0, .33)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

# Add vertical lines that show subgroup means
p1 <- p1 +
  geom_vline(xintercept = mean(subset(df, college == 1)$ideo_know, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)
p2 <- p2 +
  geom_vline(xintercept = mean(subset(df, college == 0)$ideo_know, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)

# Combine into one graph
knowledge_by_group_SI <- plot_grid(p1, p2,
                                   labels = "AUTO", ncol = 2,
                                   label_fontfamily = "Times")
knowledge_by_group_SI <-
  ggdraw(add_sub(knowledge_by_group_SI,
                 "Ideological Knowledge (0 = least knowledgeable, 4 = most knowledgeable)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
```

```
                vjust = 4.5, fontfamily = "Times"))
```

```
# Figure S4
knowledge_by_group_SI
```




```
# ggsave(file = "knowledge_by_group_SI.pdf", width = 8, height = 3)
```

```
## Conduct two-tailed t-tests for college graduates vs. non-college graduates
lm_robust(ideo_know ~ college, data = df) # reported in Appendix C
```

```
##              Estimate Std. Error  t value     Pr(>|t|)  CI Lower  CI Upper   DF
## (Intercept) 1.5495732 0.03399614 45.58085 0.000000e+00 1.4829129 1.6162335 2772
## college     0.8557026 0.05055605 16.92582 3.112599e-61 0.7565713 0.9548339 2772
```

```
## Effects of treatment on self-reported ideology by education
lm_robust(ideo ~ group * college, data = df) # reported in Appendix C
```

```
##                   Estimate Std. Error    t value   Pr(>|t|)   CI Lower
## (Intercept)     4.16379310 0.04796348 86.8117332 0.00000000  4.0697560
## group1         -0.16379310 0.08993369 -1.8212652 0.06864641 -0.3401170
## group2         -0.01730873 0.08220213 -0.2105630 0.83323973 -0.1784741
## college        -0.10539024 0.07834592 -1.3451913 0.17864503 -0.2589951
## group1:college  0.16274685 0.14362071  1.1331712 0.25721500 -0.1188357
## group2:college -0.04576703 0.13531419 -0.3382279 0.73521047 -0.3110638
##                   CI Upper   DF
## (Intercept)     4.25783024 3740
## group1          0.01253076 3740
## group2          0.14385663 3740
## college         0.04821464 3740
## group1:college  0.44432939 3740
## group2:college  0.21952977 3740
```

```
### Analysis: within-subjects differences in ideology by political knowledge ----
## Formally test the within-subject differences by political knowledge
# Politically sophisticated
lm_robust(ideo_change ~ 1,
```

```
          data = df,
          subset = (group == 2 & sophis == 1)) # reported in Appendix F
```

```
##                Estimate Std. Error  t value  Pr(>|t|)   CI Lower   CI Upper  DF
## (Intercept) -0.1140584 0.08399552 -1.35791 0.1753066 -0.2792182 0.05110147 376
```

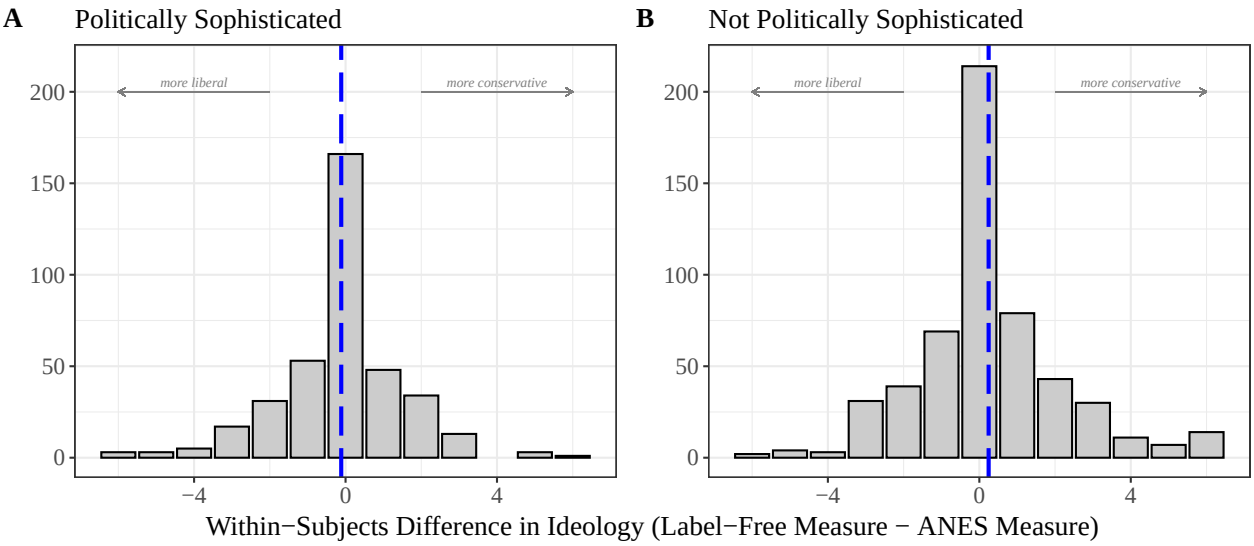
```
# Not politically sophisticated
lm_robust(ideo_change ~ 1,
          data = df,
          subset = (group == 2 & sophis == 0)) # reported in Appendix F
```

```
##                Estimate Std. Error  t value    Pr(>|t|)   CI Lower  CI Upper  DF
## (Intercept) 0.2454212 0.08334654 2.944588 0.003371849 0.08170144 0.4091411 545
```

```
## Distribution of within-subjects differences in ideology by partisanship
# Politically sophisticated
p1 <- ggplot(subset(df, group == 2 & sophis == 1), aes(x = ideo_change)) +
  geom_bar(color = "black", fill = "grey80", na.rm = T) +
  theme_bw() +
  ggtitle("Politically Sophisticated") +
  xlab("") +
  ylab("") +
  coord_cartesian(xlim = c(-6.5, 6.5), ylim = c(0, 215)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

# Not politically sophisticated
p2 <- ggplot(subset(df, group == 2 & sophis == 0), aes(x = ideo_change)) +
  geom_bar(color = "black", fill = "grey80", na.rm = T) +
  theme_bw() +
  ggtitle("Not Politically Sophisticated") +
  xlab("") +
  ylab("") +
  coord_cartesian(xlim = c(-6.5, 6.5), ylim = c(0, 215)) +
  theme(text = element_text(family = "Times", size = 12),
        axis.text = element_text(family = "Times", size = 12),
        legend.position = "none")

## Add vertical lines that show subgroup means
p1 <- p1 +
  geom_vline(xintercept = mean(subset(df, group == 2 & sophis == 1)$ideo_change, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)
p2 <- p2 +
  geom_vline(xintercept = mean(subset(df, group == 2 & sophis == 0)$ideo_change, na.rm = T),
             linetype = "longdash", color = "blue", size = 1)

## Add annotations (left arrow = more liberal; right arrow = more conservative)
p1 <- p1 +
```

```
  geom_segment(aes(x = -2, y = 200, xend = -6, yend = 200),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = -4, y = 205, label = "more liberal",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times") +
  geom_segment(aes(x = 2, y = 200, xend = 6, yend = 200),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = 4, y = 205, label = "more conservative",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times")
p2 <- p2 +
  geom_segment(aes(x = -2, y = 200, xend = -6, yend = 200),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = -4, y = 205, label = "more liberal",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times") +
  geom_segment(aes(x = 2, y = 200, xend = 6, yend = 200),
               arrow = arrow(length = unit(0.15, "cm")),
               size = .25, color = "grey50") +
  annotate("text", x = 4, y = 205, label = "more conservative",
           size = 2.5, color = "grey50", fontface = "italic", family = "Times")

## Combine into one graph
ideo_change_pid_SI <- plot_grid(p1, p2, labels = "AUTO", ncol = 2,
                                label_fontfamily = "Times")
ideo_change_pid_SI <-
  ggdraw(add_sub(ideo_change_pid_SI,
                 "Within-Subjects Difference in Ideology (Label-Free Measure - ANES Measure)",
                 vpadding = grid::unit(0, "lines"), y = 6, x = .52,
                 vjust = 4.5, fontfamily = "Times"))
```

```
# Figure S22
ideo_change_pid_SI
```




```
# ggsave(file = "ideo_change_pid_SI.pdf", width = 9, height = 4)
```

```
### Analysis: covariate balance across experimental groups (Appendix B) ----
## Create functions using code from Hartman and Hidalgo (2018)
equiv.t.test <- function(x, y, alpha = .05, epsilon = .2, std.err = "nominal",
                         cluster.x = NULL, cluster.y = NULL) {
  x = x[!is.na(x)]
  y = y[!is.na(y)]

  dbar <- mean(x) - mean(y)
  m <- as.double(length(x))
  n <- as.double(length(y))
  N <- m + n
  x.var <-  var(x)
  y.var <- var(y)
  non.cent <- (m * n * epsilon^2) / N
  critical.const <- sqrt(qf(alpha, 1, N - 2, non.cent))

  se = sqrt((m - 1) * x.var + (n - 1) * y.var) / sqrt(m * n * (N - 2) / N)
  df = N - 2

  t.stat <- dbar / se
  p = pf(abs(t.stat)^2, 1, df, non.cent)
  obs_smd = (mean(x) - mean(y)) / sd(y)
  inverted <- try(uniroot(function(x) pf(abs(t.stat)^2, 1, N - 2,
                                         ncp = (m * n * x ^ 2) / N) -
                            ifelse(pf(abs(t.stat)^2, 1, N - 2,
                                      ncp = (m * n * 0 ^ 2) / N) < alpha,
                                   pf(abs(t.stat)^2, 1, N - 2,
                                      ncp = (m * n * obs_smd^2) / N), alpha),
                          c(0, 10 * abs(t.stat)), tol = 0.0001)$root,
                  silent = TRUE)
```

```
  if(class(inverted) == "try-error") {
    inverted = NA
  }
  rej = abs(t.stat) <= critical.const
  return(list(t.stat = t.stat, critical.const = critical.const,
              power = 2 * pt(critical.const, N - 2) - 1,
              rej = rej, p = p, inverted = inverted))
}

run_equiv <- function(X, Tr, epsilon.method = "std.effect", Y = NULL,
                      custom.epsilon = NULL, std.err = "nominal",
                      type = "equiv.t.test", fdr_correct = FALSE,
                      cluster.x = NULL, cluster.y = NULL) {
  switch(epsilon.method,
         std.effect = {
           tol = abs(mean(Y[Tr == 1], na.rm = TRUE) -
                       mean(Y[Tr == 0], na.rm = TRUE)) /
             sd(Y[Tr == 0], na.rm = TRUE)
         },
         custom = {
           if(is.null(custom.epsilon)) stop("ERROR")
           tol = custom.epsilon
         },
         strict = {
           tol = 0.36
         },
         liberal = {
           tol = 0.74
         },
         stop("ERROR")
  )

  ranges = rep(tol, ncol(X))
  names(ranges) = names(X)
  print(ranges)

  # Conduct equivalence test for each variable
  tests = do.call("rbind", lapply(names(X), function(var) {
    print(var)

    res = equiv.t.test(X[Tr == 1, var], X[Tr == 0, var],
                       epsilon = ranges[var], std.err = std.err,
                       cluster.x = cluster.x, cluster.y = cluster.y)
    res$stat = res$t.stat
    res$obs_diff = (mean(X[Tr == 1, var], na.rm = TRUE) -
                      mean(X[Tr == 0, var], na.rm = TRUE))
    res$obs_smd = ((mean(X[Tr == 1, var], na.rm = TRUE) -
```

```
                        mean(X[Tr == 0, var], na.rm = TRUE))) /
      sd(X[Tr == 0, var], na.rm = TRUE)
    res$sd = sd(X[Tr == 0, var], na.rm = TRUE)

    return(c(res$inverted
             , round(res$p, 4)
             , res$obs_smd
             , res$obs_diff
             , res$power
             , res$sd))

  }))

  p.vals = unlist(tests[, 2])
  names(p.vals) = names(ranges)

  power = unlist(tests[, 5])
  names(power) = names(ranges)

  inverted = unlist(tests[, 1])
  names(inverted) = names(ranges)

  sd = unlist(tests[, 6])
  names(sd) = names(ranges)

  # Return to scale of var
  inverted.scaled = unlist(lapply(names(inverted), function(var) {
    return(inverted[var] * sd[var])
  }))
  names(inverted.scaled) = names(ranges)

  observed.smd = unlist(tests[, 3])
  names(observed.smd) = names(ranges)

  observed.diff = unlist(tests[, 4])
  names(observed.diff) = names(ranges)

  # Conduct BH FDR adjustment
  if(fdr_correct) {
    p.vals = p.adjust(p.vals, method = "BH")
  }

  return(list(tol = ranges, inverted = inverted,
              inverted.scaled = inverted.scaled, p.vals = p.vals,
              observed.diff = observed.diff, observed.smd = observed.smd,
              power = power))
}
```

```
generate_plot <- function(equiv_tests, panel.widths = c(1, 1, 5, 1, 1),
                          display.names = NULL, var.rounding = 1,
                          pval.rounding = 2, fdr_correct = FALSE) {
  .e <- environment()
  if(!is.null(display.names)) {
    equiv_tests$names = factor(display.names, levels = rev(display.names))
  } else {
    equiv_tests$names = factor(row.names(equiv_tests),
                               levels = rev(row.names(equiv_tests)))
  }
  equiv_tests$const = rep(1, nrow(equiv_tests))
  equiv_tests = equiv_tests[nrow(equiv_tests):1, ]

  g = ggplot(equiv_tests, aes(x = names))
  print(length(unique(equiv_tests$tol)))
  g = g + geom_linerange(aes(ymin = -1 * inverted, ymax = inverted),
                         size = 5, color = "darkgray", alpha = 0.9)
  if(length(unique(equiv_tests$tol)) > 1) {
    g = g + geom_linerange(aes(ymin = -1 * tol, ymax = tol),
                           size = 10, color = "gray")
  } else {
    print("here")
    print(unique(equiv_tests$tol))
    lines = unique(equiv_tests$tol)
    g = g + geom_hline(yintercept = c(-1 * lines, lines),
                       linetype = 2, size = 0.75)
  }
  g = g + scale_shape_identity() +
    geom_point(aes(y = observed.smd), color = "black", shape = 18, size = 4)
  g = (g + coord_flip() + theme(
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    axis.title.y = element_blank(),
    axis.text.x = element_text(size = 14),
    axis.title.x = element_text(size = 14)) +
      labs(x = NULL,
           y = paste("Equivalence Range (in pooled SD)"),
           font = 5) +
      theme_bw()
  )
  if(length(unique(equiv_tests$tol)) == 1) {
    g = g + ggtitle(paste0("Equivalence Tests  \n \n", "Equivalence Range: +/- ",
                           round(unique(equiv_tests$tol), 2), " pooled SD\n"))
  } else {
    g = g + ggtitle("Equivalence Tests  \n \n \n")
  }
```

```
g_inv = ggplot(equiv_tests, aes(x = names, y = const,
                                label = round(inverted.scaled, var.rounding)),
               environment = .e)
g_inv = g_inv + geom_text()
g_inv = (g_inv + theme_bw() + coord_flip() +
           theme(panel.grid.minor = element_blank(),
                 panel.grid.major=element_blank(),
                 axis.text.x = element_text(color = "white", size = 14),
                 axis.ticks.x = element_line(color = "white"),
                 axis.text.y = element_blank(),
                 axis.ticks.y = element_blank(),
                 axis.title.x = element_text(size = 14)) +
           ylim(1 - .05, 1.05) +
           ggtitle("Equivalence\nConfidence\nInterval (+/-)\n(Scale of Var)") +
           labs(y = " ", x = NULL)
)

g_obs = ggplot(equiv_tests, aes(x = names, y = const,
                                label = round(observed.diff, var.rounding)),
               environment = .e)
g_obs = g_obs + geom_text()
g_obs = (g_obs + theme_bw() + coord_flip() +
           theme(panel.grid.minor = element_blank(),
                 panel.grid.major = element_blank(),
                 axis.text.x = element_text(color = "white", size = 14),
                 axis.ticks.x = element_line(color = "white"),
                 axis.text.y = element_blank(),
                 axis.ticks.y = element_blank(),
                 axis.title.x = element_text(size = 14)) +
           ylim(1 - .05, 1.05) +
           ggtitle("Observed\nMean\nDifference\n(Scale of Var)") +
           labs(y = " ", x = NULL)
)

g_pval = ggplot(equiv_tests, aes(x = names, y = const,
                                 label = round(p.vals, pval.rounding)),
                environment = .e)
g_pval = g_pval + geom_text()
g_pval = (g_pval + theme_bw() + coord_flip() +
            theme(panel.grid.minor = element_blank(),
                  panel.grid.major = element_blank(),
                  axis.text.x = element_text(color = "white", size = 14),
                  axis.ticks.x = element_line(color = "white"),
                  axis.text.y = element_blank(),
                  axis.ticks.y = element_blank(),
                  axis.title.x = element_text(size = 14)) +
            ylim(1 - .05, 1.05) +
```

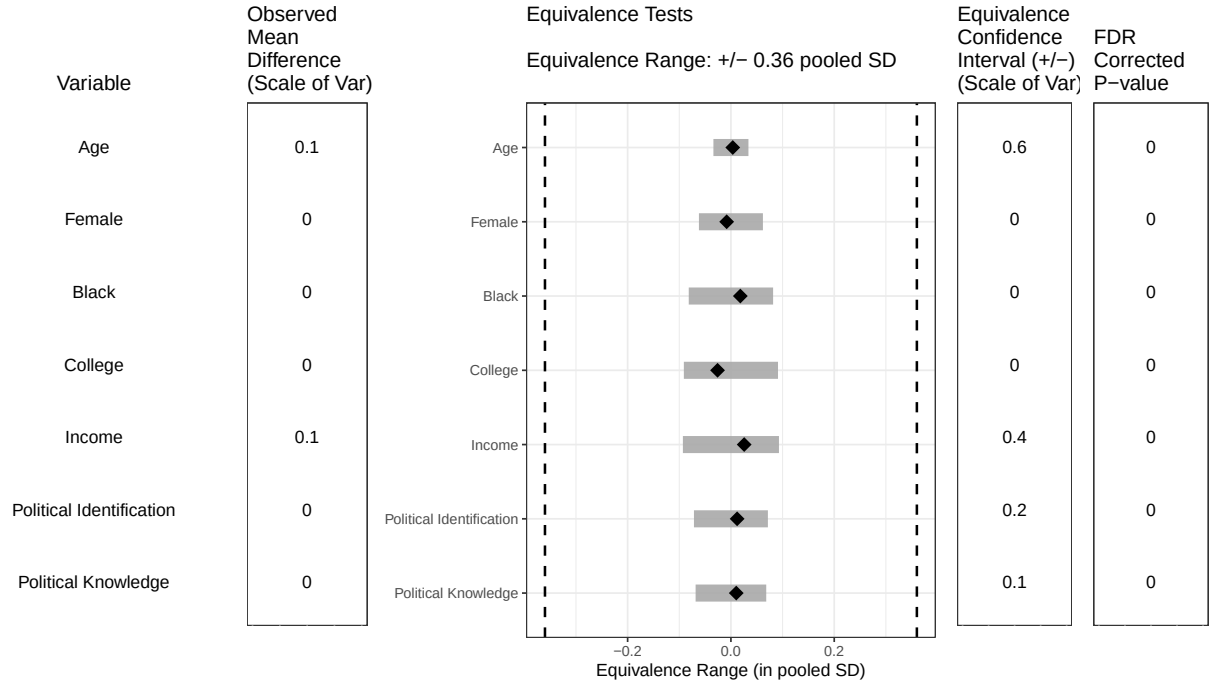
```
              ggtitle(ifelse(fdr_correct, "\nFDR\nCorrected\nP-value",
                             "\n\n\nP-value")) +
              labs(y = " ", x = NULL)
  )

  g_var = ggplot(equiv_tests, aes(x = names, y = const, label = names))
  g_var = g_var + geom_text()
  g_var = (g_var + theme_bw() + coord_flip() +
             theme(panel.grid.minor = element_blank(),
                   panel.grid.major=element_blank(),
                   axis.text.x = element_text(color = "white", size = 14),
                   axis.ticks.x = element_line(color = "white"),
                   axis.text.y = element_blank(),
                   axis.ticks.y = element_blank(),
                   panel.border = element_blank(),
                   axis.title.x = element_text(size = 14)) +
             ylim(1-.05, 1.05) +
             ggtitle("\n \n \nVariable") +
             theme(plot.title = element_text(hjust = 0.5)) +
             labs(y = " ", x = NULL)
  )

  grid.arrange(g_var, g_obs, g, g_inv, g_pval, ncol = 5, nrow = 1,
               widths = panel.widths, heights = c(4))
}

## Equivalence tests for respondents in control group and Add Definitions condition
replicate_strict <- as.data.frame(
  run_equiv(X = subset(df,
                       select = c("age", "female", "black", "college",
                                  "income", "pid", "pol_know")),
            Tr = df$group, epsilon.method = "custom", custom.epsilon = 0.36,
            fdr_correct = TRUE)
)
```

```
##      age   female    black  college   income      pid pol_know
##     0.36     0.36     0.36     0.36     0.36     0.36     0.36
## [1] "age"
## [1] "female"
## [1] "black"
## [1] "college"
## [1] "income"
## [1] "pid"
## [1] "pol_know"
```

```
p <-
  generate_plot(replicate_strict, panel.widths = c(2.5, 1.25, 5, 1.2, 1.2),
                display.names = c("Age", "Female", "Black", "College", "Income",
```

```
                                    "Political Identification", "Political Knowledge"),
                pval.rounding = 3, fdr_correct = TRUE)
```

```
## [1] 1
## [1] "here"
## [1] 0.36
```




```
# ggsave(file = "balance_C_T1.pdf", width = 11, height = 6)
```

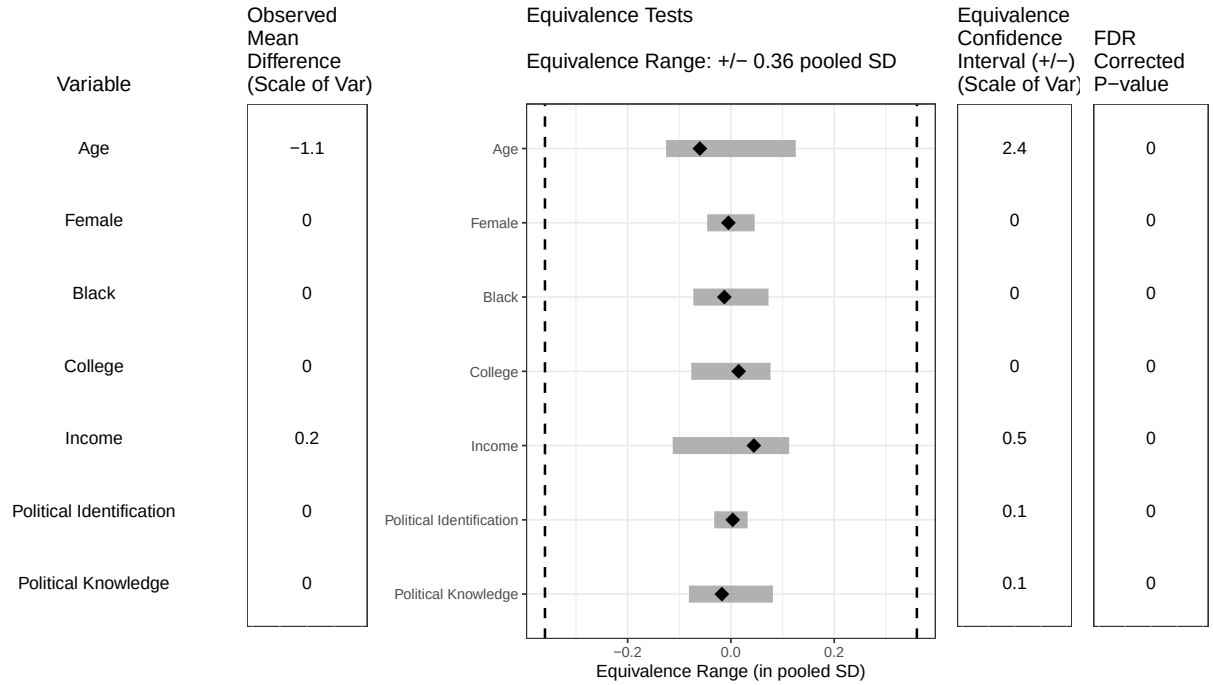
```
## Equivalence tests for respondents in control group and Subtract Labels condition
df <- df %>% mutate(group_alt1 = case_when(
  group == 0 ~ 0,
  group == 1 ~ 2,
  group == 2 ~ 1))

replicate_strict <- as.data.frame(
  run_equiv(X = subset(df,
                       select = c("age", "female", "black", "college",
                                  "income", "pid", "pol_know")),
            Tr = df$group_alt1, epsilon.method = "custom", custom.epsilon = 0.36,
            fdr_correct = TRUE)
)
```

```
##      age   female    black  college   income      pid pol_know
##     0.36     0.36     0.36     0.36     0.36     0.36     0.36
## [1] "age"
## [1] "female"
## [1] "black"
```

```
## [1] "college"
## [1] "income"
## [1] "pid"
## [1] "pol_know"
```

```
p <-
  generate_plot(replicate_strict, panel.widths = c(2.5, 1.25, 5, 1.2, 1.2),
                display.names = c("Age", "Female", "Black", "College", "Income",
                                  "Political Identification", "Political Knowledge"),
                pval.rounding = 3, fdr_correct = TRUE)
```

```
## [1] 1
## [1] "here"
## [1] 0.36
```




```
# ggsave(file = "balance_C_T2.pdf", width = 11, height = 6)
```

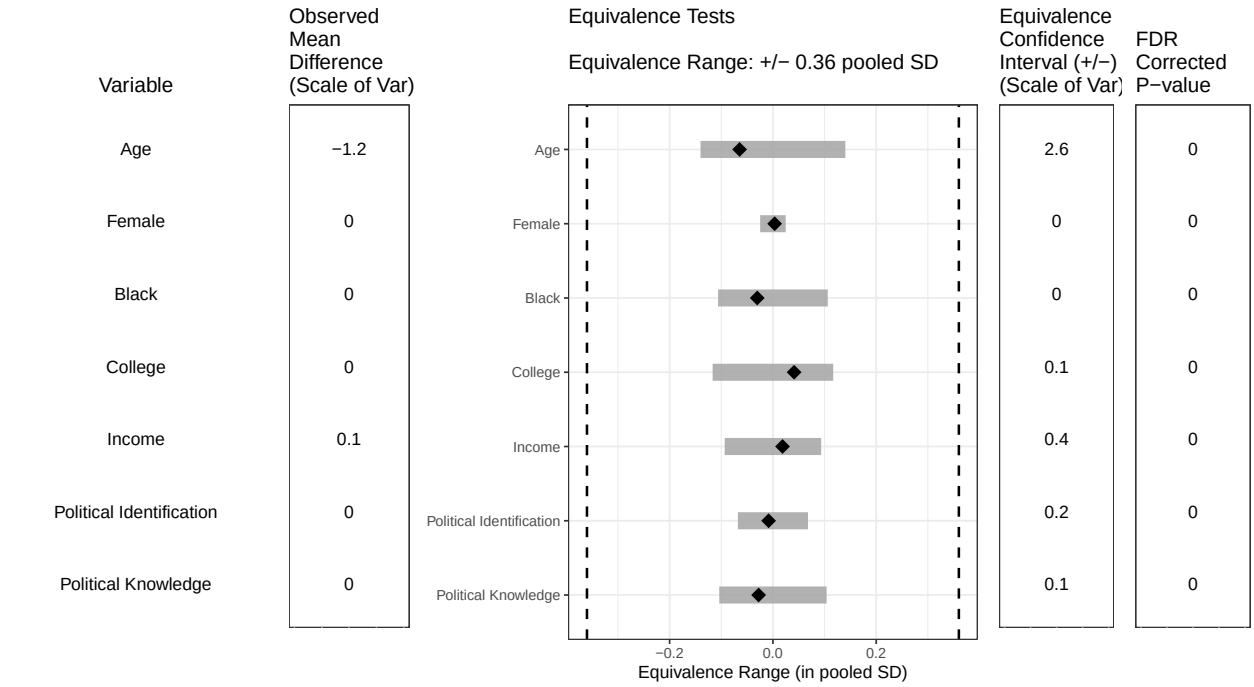
```
## Equivalence tests for respondents in Add Definitions and Subtract Labels conditions
df <- df %>% mutate(group_alt2 = case_when(
  group == 0 ~ 2,
  group == 1 ~ 0,
  group == 2 ~ 1))

replicate_strict <- as.data.frame(
  run_equiv(X = subset(df,
                       select = c("age", "female", "black", "college",
                                  "income", "pid", "pol_know")),
            Tr = df$group_alt2, epsilon.method = "custom", custom.epsilon = 0.36,
            fdr_correct = TRUE)
```

```
)
```

```
##      age   female    black  college   income      pid pol_know
##     0.36     0.36     0.36     0.36     0.36     0.36     0.36
## [1] "age"
## [1] "female"
## [1] "black"
## [1] "college"
## [1] "income"
## [1] "pid"
## [1] "pol_know"
```

```
p <- generate_plot(replicate_strict, panel.widths = c(2.5, 1.25, 5, 1.2, 1.2),
                   display.names = c("Age", "Female", "Black", "College", "Income",
                                     "Political Identification", "Political Knowledge"),
                   pval.rounding = 3, fdr_correct = TRUE)
```

```
## [1] 1
## [1] "here"
## [1] 0.36
```




```
# ggsave(file = "balance_T1_T2.pdf", width = 11, height = 6)
```

```
### Analysis: word counting for open-ended responses (Appendix C.5) ----
## Subset the dataset
# Self-reported liberals in control group
df_lib <- df %>%
  filter(randomizer == 1 | randomizer == 2) %>%
  filter(liberal == 1) %>%
```

```
  filter(lib_open != "")

# Self-reported conservatives in control group
df_con <- df %>%
  filter(randomizer == 1 | randomizer == 2) %>%
  filter(conserv == 1) %>%
  filter(cons_open != "")

## Create a dictionary
myDict <-
  dictionary(list(dem = c("Democrat", "Democrats", "Democratic Party"),
                  gop = c("Republican", "Republicans", "Republican Party"),
                  ideo = c("activ*", "chang*", "tradition*", "interven*")))

## Analysis for df_lib
# Tokenization and normalization
dtm_lib <- dfm(df_lib$lib_open,           # input text
               tolower = T, stem = T)     # lowercase and stem
```

```
## Warning: 'dfm.character()' is deprecated. Use 'tokens()' first.
```

```
## Warning: 'stem' is deprecated; use dfm_wordstem() instead
```

```
# Analyze the open-ended responses based on simple counting
dtm_lib <- dfm_lookup(dtm_lib, myDict)
```

```
## as(<dgCMatrix>, "dgTMatrix") is deprecated since Matrix 1.5-0; do as(., "TsparseMatrix") ins
```

```
dtm_lib <- convert(dtm_lib, to = "data.frame") # convert to data frame
dtm_lib$dem <- ifelse(dtm_lib$dem > 0, 1, 0)
dtm_lib$gop <- ifelse(dtm_lib$gop > 0, 1, 0)
dtm_lib$ideo <- ifelse(dtm_lib$ideo > 0, 1, 0)
sum(dtm_lib$dem)  # this counts how many responses contain Democratic vocab
```

```
## [1] 17
```

```
sum(dtm_lib$gop)  # this counts how many responses contain Republican vocab
```

```
## [1] 5
```

```
sum(dtm_lib$ideo) # this counts how many responses contain ideological vocab
```

```
## [1] 36
```

```
## Analysis for df_con
# Tokenization and normalization
dtm_con <- dfm(df_con$cons_open,           # input text
               tolower = T, stem = T)      # lowercase and stem
```

```
## Warning: 'dfm.character()' is deprecated. Use 'tokens()' first.
```

```
## Warning: 'stem' is deprecated; use dfm_wordstem() instead
```

```r
# Analyze the open-ended responses based on simple counting
dtm_con <- dfm_lookup(dtm_con, myDict)
dtm_con <- convert(dtm_con, to = "data.frame") # convert to data frame
dtm_con$dem <- ifelse(dtm_con$dem > 0, 1, 0)
dtm_con$gop <- ifelse(dtm_con$gop > 0, 1, 0)
dtm_con$ideo <- ifelse(dtm_con$ideo > 0, 1, 0)
sum(dtm_con$dem)  # this counts how many responses contain Democratic vocab
```

```
## [1] 8
```

```r
sum(dtm_con$gop)  # this counts how many responses contain Republican vocab
```

```
## [1] 22
```

```r
sum(dtm_con$ideo) # this counts how many responses contain ideological vocab
```

```
## [1] 40
```