Main_Replication_Code.R

anton

Fri Oct 20 15:20:01 2017

```
############ Replication Code - "Investigator Characteristics and Respondent Behavior in Online Surveys
############

#### Clear workspace
rm(list=ls())

#### Libraries
library(foreign)
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(ggplot2)
library(digest)
library(xtable)

# colour scheme
theme_bw1 <- function(base_size = 16, base_family = "") {
  theme_bw(base_size = base_size, base_family = base_family) %+replace%
    theme(
      axis.text.x =       element_text(size = base_size*.9, colour = "black"),
      axis.text.y =       element_text(size = base_size, colour = "black"),
      axis.ticks =        element_line(colour = "grey75"),
      axis.title.x =      element_text(size=base_size, face="bold"),
      axis.title.y =      element_text(size = base_size,angle=90, face="bold"),
      plot.title =        element_text(face = "bold"),
      legend.position =   "none"
    )
}

#### Load dataset
researcher.data.full <- read.csv("data_investigator_characteristics.csv", stringsAsFactors=F)

###########################
#### This code was run but omitted from replication as the distributed replication file already contains
#### Hash MTurk IDs
#researcher.data.full$ID1Hash <- sapply(researcher.data.full$mTurkID, function(x) digest(x, "md5"))
#researcher.data.full$ID1Hash[researcher.data.full$mTurkID == ""] <- NA ###  Missing MTurk IDs
#researcher.data.full$ID2Hash <- sapply(researcher.data.full$IP, function(x) digest(x, "md5"))

### Convert to arbitrary numeric identifiers
```

```
#researcher.data.full$ID1Number <- as.integer(as.factor(researcher.data.full$ID1Hash))
#researcher.data.full$ID2Number <- as.integer(as.factor(researcher.data.full$ID2Hash))
############################
############################


###Remove duplicated ID's, as described in SI Section B
### Starting time
researcher.data.full$startTime <- strptime(as.character(researcher.data.full$StartDate), format = "%m/%

### Sort by startTime
researcher.data.full <- researcher.data.full[order(researcher.data.full$startTime),]

### Is Duplicated by ID1Hash?
researcher.data.full$ID1Duplicated <- NA
researcher.data.full$ID1Duplicated[!is.na(researcher.data.full$ID1Number)] <- as.integer(duplicated(res
### Users with ID1Number #548 and #979 appears twice by IP, have MTurk ID only in the chronologically s
researcher.data.full$ID1Duplicated[researcher.data.full$ID1Number == 548|researcher.data.full$ID1Number

### For those with null ID1 hash, are they duplicated by IP
ID2Temp <- duplicated(researcher.data.full$ID2Number)
researcher.data.full$ID2Duplicated <- 0
researcher.data.full$ID2Duplicated[is.na(researcher.data.full$ID1Duplicated)] <- ID2Temp[is.na(researchc

### Merge into a single "duplicated" vector
researcher.data.full$duplicated <- as.integer(researcher.data.full$ID1Duplicated)
researcher.data.full$duplicated[is.na(researcher.data.full$ID1Duplicated)] <- as.integer(researcher.data

### Drop duplicated observations from indviduals taking the survey multiple times.
researcher.data.full <- subset(researcher.data.full, duplicated != 1)

### How many observations are left?
nrow(researcher.data.full)
```

```
## [1] 2036
```

```
########## Get treatments
### Convert names to treatment labels
blackwomen <- c('Ebony Gaines', 'Deja Washington')
whitemen <- c('Brett Walsh', 'Connor Schroeder')
blackmen <- c('DeShawn Booker', 'Tyrone Robinson')
whitewomen <- c('Laurie Yoder', 'Molly Ryan')

# Treatments
researcher.data.full$treatment.whitewoman <- ifelse(researcher.data.full$researcherName %in% whitewomen
researcher.data.full$treatment.blackwoman <- ifelse(researcher.data.full$researcherName %in% blackwomen
researcher.data.full$treatment.whiteman <- ifelse(researcher.data.full$researcherName %in% whitemen, 1,
researcher.data.full$treatment.blackman <- ifelse(researcher.data.full$researcherName %in% blackmen, 1,

# Woman/Black treatments
researcher.data.full$treatment.woman <- ifelse(researcher.data.full$treatment.whitewoman == 1 | researcl
researcher.data.full$treatment.black <- ifelse(researcher.data.full$treatment.blackwoman == 1 | researcl

# Did you complete the survey and could we match you?
```

```r
researcher.data.full$finishedAndMatched <- as.integer(!is.na(researcher.data.full$ID1Number)&researcher

# How many finished and matched
sum(researcher.data.full$finishedAndMatched)
```

```
## [1] 2005
```

```r
############################
### More data setup
############################

################################
#### Subset data to units that finished
researcher.data <- subset(researcher.data.full, finishedAndMatched  == 1)
#### All -99 are NAs
researcher.data[researcher.data==-99] <- NA


#####################################################
### Pass attention checks?

researcher.data$ac1.pass <- ifelse(researcher.data$ac1.huffpo == 1 & researcher.data$ac1.Reuters == 1,
researcher.data$ac2.pass <- ifelse(researcher.data$ac2.Energy == 1 & researcher.data$ac2.GlobalTrade ==

researcher.data$ac1.pass[is.na(researcher.data$ac1.pass)] <- 0
researcher.data$ac2.pass[is.na(researcher.data$ac2.pass)] <- 0

researcher.data$ac.pass.both <- ifelse(researcher.data$ac1.pass== 1 & researcher.data$ac2.pass == 1, 1,

#####################################################
### Background covariates

researcher.data$gender.male <- ifelse(researcher.data$gender == 1, 1, 0)
researcher.data$gender.female <- ifelse(researcher.data$gender == 2, 1, 0)

researcher.data$democrat <- ifelse(researcher.data$party == 1, 1, 0)
researcher.data$republican <- ifelse(researcher.data$party ==2, 1, 0)
researcher.data$independent <- ifelse(researcher.data$party == 3, 1, 0)

researcher.data$voteyes <- ifelse(researcher.data$turnout == 3, 1, 0)
researcher.data$voteno <- ifelse(researcher.data$turnout == 2, 1, 0)
researcher.data$votedk <- ifelse(researcher.data$turnout == 1, 1, 0)

researcher.data$white <- ifelse(researcher.data$ethnicity.white == 1, 1, 0)
researcher.data$black <- ifelse(researcher.data$ethnicity.black == 1, 1, 0)


################################################

###### Racial_resentment Scale
# Convert to Factor
# recall that items 1/4 are reverse-coded compared to 2/3.

researcher.data$blacks.less.than.deserve[researcher.data$blacks.less.than.deserve == -99] <- NA
```

```
researcher.data$blacks.less.than.deserve <- factor(researcher.data$blacks.less.than.deserve,
                                                   labels=c("Strongly Agree", "Agree", "Disagree", "Str
researcher.data$blacks.less.than.deserve <- factor(researcher.data$blacks.less.than.deserve,levels(resea

researcher.data$blacks.like.other.minorities[researcher.data$blacks.like.other.minorities == -99] <- NA
researcher.data$blacks.like.other.minorities <- factor(researcher.data$blacks.like.other.minorities,
                                                       labels=c("Strongly Agree", "Agree", "Disagree",
researcher.data$blacks.like.other.minorities <- factor(researcher.data$blacks.like.other.minorities,leve

researcher.data$blacks.not.trying[researcher.data$blacks.not.trying == -99] <- NA
researcher.data$blacks.not.trying <- factor(researcher.data$blacks.not.trying,
                                            labels=c("Strongly Agree", "Agree", "Disagree", "Strongly D
researcher.data$blacks.not.trying <- factor(researcher.data$blacks.not.trying,levels(researcher.data$bla

researcher.data$blacks.slavery.discrimination[researcher.data$blacks.slavery.discrimination == -99] <- N
researcher.data$blacks.slavery.discrimination <- factor(researcher.data$blacks.slavery.discrimination,
                                                        labels=c("Strongly Agree", "Agree", "Disagree",

researcher.data$blacks.slavery.discrimination <- factor(researcher.data$blacks.slavery.discrimination,le


researcher.data$racial.resentment <- ((as.numeric(researcher.data$blacks.less.than.deserve)-1)/4 + (as.n
                                      (as.numeric(researcher.data$blacks.not.trying)-1)/4 +  (as.numer

################################################
## Make Panel 1 of Figure 1 (Racial Resentment)
################################################

### Model frame for predictions:
prediction.frame <- data.frame(treatment.black = c(1,0))

###### Estimate effect on racial resentment
researcher.data.resentment <- subset(researcher.data, !is.na(racial.resentment))
resentment <- lm(racial.resentment  ~ treatment.black, data=researcher.data.resentment)

### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims.resentment <- matrix(ncol=length(coef(resentment)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.resentment[sample(1:nrow(researcher.data.resentment), nrow(res
  researcher.boot <- lm(racial.resentment~ treatment.black , data=researcher.data.boot)
  boot.estims.resentment[it,] <- coef(researcher.boot)
}

resentment_vcov <- cov(boot.estims.resentment)

resentment_pval <- round(coeftest(resentment, resentment_vcov)[2,4], 3)
resentment_results <- data.frame(treatments = c(paste("Black Name\nn=",sum(model.frame(resentment)$treat
                                                paste("White Name\nn=",sum(model.frame(resentment)$treat
                                 point = predict(resentment, newdata=prediction.frame),
                                 se = sqrt(c(resentment_vcov[1,1] + resentment_vcov[2,2] + 2*resentment_
```

```
resentment_results$lower95 <- resentment_results$point - qnorm(.975)*resentment_results$se
resentment_results$upper95 <- resentment_results$point + qnorm(.975)*resentment_results$se


#### Plot racial resentment

resentment_diff <- data.frame(treatments = c(paste("Black Name (vs. White Name)\nn=",nrow(model.frame(r
           point = coef(resentment)[2],
           se = sqrt(resentment_vcov[2,2]))

resentment_diff$lower95 <- resentment_diff$point - qnorm(.975)*resentment_diff$se
resentment_diff$upper95 <- resentment_diff$point + qnorm(.975)*resentment_diff$se

#### Plot racial resentment - difference
pdf("figures/Figure1_Panel1.pdf", width=4, height=7.1)
p = ggplot(resentment_diff, aes(y=point, x=treatments))
p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on Racial Resentment Scale", limits=c(-.2, .2))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("p-value of difference = ",resentment_pval, sep=""))
p = p + ggtitle("Effect of researcher race\non racial resentment\n")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
#####################################################
## Make Panel 2 of Figure 1 (Black President)
#####################################################

######## Black president
researcher.data$president.black.yes <- ifelse(researcher.data$president.black == 1, 1, 0)
researcher.data$president.black.no <- ifelse(researcher.data$president.black == 2, 1, 0)


researcher.data.blackpres <- subset(researcher.data, !is.na(researcher.data$president.black.yes))
black_president <- lm(president.black.yes ~ treatment.black, data=researcher.data.blackpres)

### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims.blackpres <- matrix(ncol=length(coef(black_president)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.blackpres[sample(1:nrow(researcher.data.blackpres), nrow(resea
  researcher.boot <- lm(president.black.yes ~ treatment.black , data=researcher.data.boot)
  boot.estims.blackpres[it,] <- coef(researcher.boot)
}

black_president_vcov <- cov(boot.estims.blackpres)
```

```
black_president_pval <- round(coeftest(black_president, black_president_vcov)[2,4], 3)
prediction.frame <- data.frame(treatment.black = c(1,0))

black_president_results <- data.frame(treatments = c(paste("Black Name\nn=",sum(model.frame(black_presi
                                                     paste("White Name\nn=",sum(model.frame(black_presi
                                      point = predict(black_president, newdata=prediction.frame),
                                      se = sqrt(c(black_president_vcov[1,1] + black_president_vcov[2,2]

black_president_results$lower95 <- black_president_results$point - qnorm(.975)*black_president_results$
black_president_results$upper95 <- black_president_results$point + qnorm(.975)*black_president_results$


blackpres_diff <- data.frame(treatments = c(paste("Black Name (vs. White Name)\nn=",nrow(model.frame(bla
                               point = coef(black_president)[2],
                               se = sqrt(black_president_vcov[2,2])))

blackpres_diff$lower95 <- blackpres_diff$point - qnorm(.975)*blackpres_diff$se
blackpres_diff$upper95 <- blackpres_diff$point + qnorm(.975)*blackpres_diff$se

#### Plot black president - difference
pdf("figures/Figure1_Panel2.pdf", width=4, height=7.1)
p = ggplot(blackpres_diff, aes(y=point, x=treatments))

p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on share that would vote for a black president", limits=c(-.2,
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("p-value of difference = ",black_president_pval, sep=""))
p = p + ggtitle("Effect of researcher race on\nwillingness to vote for\na black president")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
################################################
## Now make Figure 2 Panel 1 (Gender Equality)
################################################

###### Women's Role
researcher.data$womenrole[researcher.data$womenrole == -99] <- NA
researcher.data$womenrole.equal <- ifelse(researcher.data$womenrole == 1, 1, 0)
researcher.data$womenrole.home <- ifelse(researcher.data$womenrole == 2, 1, 0)
researcher.data$womenrole.dk <- ifelse(researcher.data$womenrole == 3, 1, 0)


researcher.data.womenequal <- subset(researcher.data, !is.na(researcher.data$womenrole.equal))
women_equal <- lm(womenrole.equal ~ treatment.woman, data=researcher.data.womenequal)


### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
```

```
boot.estims.womenequal <- matrix(ncol=length(coef(women_equal)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.womenequal[sample(1:nrow(researcher.data.womenequal), nrow(res
  researcher.boot <- lm(womenrole.equal ~ treatment.woman , data=researcher.data.boot)
  boot.estims.womenequal[it,] <- coef(researcher.boot)
}

women_equal_vcov <- cov(boot.estims.womenequal)


women_equal_pval <- round(coeftest(women_equal, women_equal_vcov)[2,4], 3)
prediction.frame.woman <- data.frame(treatment.woman = c(1,0))

women_equal_results <- data.frame(treatments = c(paste("Female Name\nn=",sum(model.frame(women_equal)$t
                                                 paste("Male Name\nn=",sum(model.frame(women_equal)$trea
                                  point = predict(women_equal, newdata=prediction.frame.woman),
                                  se = sqrt(c(women_equal_vcov[1,1] + women_equal_vcov[2,2] + 2*women_e

women_equal_results$lower95 <- women_equal_results$point - qnorm(.975)*women_equal_results$se
women_equal_results$upper95 <- women_equal_results$point + qnorm(.975)*women_equal_results$se


#### Alternate figure - difference
women_equal_diff <- data.frame(treatments = c(paste("Female Name (vs. Male Name)\nn=",nrow(model.frame(
                               point = coef(women_equal)[2],
                               se = sqrt(women_equal_vcov[2,2])))

women_equal_diff$lower95 <- women_equal_diff$point - qnorm(.95)*women_equal_diff$se
women_equal_diff$upper95 <- women_equal_diff$point + qnorm(.95)*women_equal_diff$se

### BH-adjusted confidence bounds = 1 - .05 * (1/3)% bound
women_equal_diff$lowerBH <-  women_equal_diff$point - qnorm(1 - .05 * (1/3))*women_equal_diff$se
women_equal_diff$upperBH <-  women_equal_diff$point + qnorm(1 - .05 * (1/3))*women_equal_diff$se

#### Plot women equal
pdf("figures/Figure2_Panel1.pdf", width=4, height=7.1)
p = ggplot(women_equal_diff, aes(y=point, x=treatments))

p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on share that say women should be equal to men", limits=c(-.2,
p = p + geom_pointrange(aes(ymin=lowerBH,ymax=upperBH), size= 1.5)

p = p + scale_x_discrete(name=paste("p-value of difference = ",women_equal_pval, sep=""))
p = p + ggtitle("Effect of researcher gender on\nattitudes on gender equality\n")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
##########################################
## Make Figure 2 Panel 2 (Woman President)
```

```
########################################

######## Woman president
researcher.data$president.woman.yes <- ifelse(researcher.data$president.woman == 1, 1, 0)
researcher.data$president.woman.no <- ifelse(researcher.data$president.woman == 2, 1, 0)

researcher.data.womanpres <- subset(researcher.data, !is.na(researcher.data$president.woman.yes))
woman_president <- lm(president.woman.yes ~ treatment.woman, data=researcher.data.womanpres)

### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims.womanpres <- matrix(ncol=length(coef(woman_president)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.womanpres[sample(1:nrow(researcher.data.womanpres), nrow(resea
  researcher.boot <- lm(president.woman.yes ~ treatment.woman , data=researcher.data.boot)
  boot.estims.womanpres[it,] <- coef(researcher.boot)
}

woman_president_vcov <- cov(boot.estims.womanpres)

# Marginals
woman_president_pval <- round(coeftest(woman_president, woman_president_vcov)[2,4], 3)
prediction.frame.woman <- data.frame(treatment.woman = c(1,0))

woman_president_results <- data.frame(treatments = c(paste("Female Name\nn=",sum(model.frame(woman_presi
                                                     paste("Male Name\nn=",sum(model.frame(woman_preside
                                      point = predict(woman_president, newdata=prediction.frame.woman),
                                      se = sqrt(c(woman_president_vcov[1,1] + woman_president_vcov[2,2]

woman_president_results$lower95 <- woman_president_results$point - qnorm(.975)*woman_president_results$s
woman_president_results$upper95 <- woman_president_results$point + qnorm(.975)*woman_president_results$s


#### difference
woman_president_diff <- data.frame(treatments = c(paste("Female Name (vs. Male Name)\nn=",nrow(model.fra
                                point = coef(woman_president)[2],
                                se = sqrt(woman_president_vcov[2,2])))

woman_president_diff$lower95 <- woman_president_diff$point - qnorm(.975)*woman_president_diff$se
woman_president_diff$upper95 <- woman_president_diff$point + qnorm(.975)*woman_president_diff$se



#### Plot woman president - difference
pdf("figures/Figure2_Panel2.pdf", width=4, height=7.1)
p = ggplot(woman_president_diff, aes(y=point, x=treatments))

p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on share that would vote for a woman president", limits=c(-.2,
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)
```

```
p = p + scale_x_discrete(name=paste("p-value of difference = ",woman_president_pval, sep=""))
p = p + ggtitle("Effect of researcher gender on\nwillingness to vote for\na woman president")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
################################################
## Make Figure 1 Panel 3 and Figure 2 Panel 3
################################################

#### Service spending : both race- and gender-of-interviewer effects
researcher.data$services.more <- ifelse(researcher.data$services == 2, 1, 0)
researcher.data$services.cut <- ifelse(researcher.data$services == 1, 1, 0)

researcher.data.servicesmore <- subset(researcher.data, !is.na(researcher.data$services.more))

service_cut_black <- lm(services.more ~ treatment.black, data=researcher.data.servicesmore)


### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims.servicesmoreblack <- matrix(ncol=length(coef(service_cut_black)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.servicesmore[sample(1:nrow(researcher.data.servicesmore), nrow
  researcher.boot <- lm(services.more ~ treatment.black, data=researcher.data.boot)
  boot.estims.servicesmoreblack[it,] <- coef(researcher.boot)
}

service_cut_black_vcov <- cov(boot.estims.servicesmoreblack)


# Marginals
service_cut_black_pval <- round(coeftest(service_cut_black, service_cut_black_vcov)[2,4], 3)
prediction.frame.black <- data.frame(treatment.black = c(1,0))

service_cut_black_results <- data.frame(treatments = c(paste("Black Name\nn=",sum(model.frame(service_cu
                                                       paste("White Name\nn=",sum(model.frame(service_cu
                                        point = predict(service_cut_black, newdata=prediction.frame.blac
                                        se = sqrt(c(service_cut_black_vcov[1,1] + service_cut_black_vcov

service_cut_black_results$lower95 <- service_cut_black_results$point - qnorm(.975)*service_cut_black_res
service_cut_black_results$upper95 <- service_cut_black_results$point + qnorm(.975)*service_cut_black_res


service_cut_black_diff <- data.frame(treatments = c(paste("Black Name (vs. White Name)\nn=",nrow(model.
                                    point = coef(service_cut_black)[2],
                                    se = sqrt(service_cut_black_vcov[2,2])))

service_cut_black_diff$lower95 <- service_cut_black_diff$point - qnorm(.975)*service_cut_black_diff$se
```

```
service_cut_black_diff$upper95 <- service_cut_black_diff$point + qnorm(.975)*service_cut_black_diff$se

#### Plot service cut black - diff
pdf("figures/Figure1_Panel3.pdf", width=4, height=7.1)
p = ggplot(service_cut_black_diff, aes(y=point, x=treatments))

p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on share supporting greater service spending", limits=c(-.2, .2)
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("p-value of difference = ",service_cut_black_pval, sep=""))
p = p + ggtitle("Effect of researcher race on\nwillingness to support\npublic service expansion")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
### Gender effect

service_cut_woman <- lm(services.more ~ treatment.woman, data=researcher.data.servicesmore)


### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims.servicesmorewoman <- matrix(ncol=length(coef(service_cut_woman)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.servicesmore[sample(1:nrow(researcher.data.servicesmore), nrow
  researcher.boot <- lm(services.more ~ treatment.woman, data=researcher.data.boot)
  boot.estims.servicesmorewoman[it,] <- coef(researcher.boot)
}

service_cut_woman_vcov <- cov(boot.estims.servicesmorewoman)

service_cut_woman_pval <- round(coeftest(service_cut_woman, service_cut_woman_vcov)[2,4], 3)
prediction.frame.woman <- data.frame(treatment.woman = c(1,0))

service_cut_woman_results <- data.frame(treatments = c(paste("Female Name\nn=",sum(model.frame(service_
                                                       paste("Male Name\nn=",sum(model.frame(service_cu
                                        point = predict(service_cut_woman, newdata=prediction.frame.woma
                                        se = sqrt(c(service_cut_woman_vcov[1,1] + service_cut_woman_vcov

service_cut_woman_results$lower95 <- service_cut_woman_results$point - qnorm(.975)*service_cut_woman_res
service_cut_woman_results$upper95 <- service_cut_woman_results$point + qnorm(.975)*service_cut_woman_res


#### Plot service cut woman

service_cut_woman_diff <- data.frame(treatments = c(paste("Female Name (vs. Male Name)\nn=",nrow(model.
                                    point = coef(service_cut_woman)[2],
                                    se = sqrt(service_cut_woman_vcov[2,2]))
```

```
service_cut_woman_diff$lower95 <- service_cut_woman_diff$point - qnorm(.975)*service_cut_woman_diff$se
service_cut_woman_diff$upper95 <- service_cut_woman_diff$point + qnorm(.975)*service_cut_woman_diff$se

### BH-adjusted confidence bounds = 1 - .05 * (2/3)% bound
service_cut_woman_diff$lowerBH <- service_cut_woman_diff$point - qnorm(.9833333)*service_cut_woman_diff$
service_cut_woman_diff$upperBH <- service_cut_woman_diff$point + qnorm(.9833333)*service_cut_woman_diff$

#### Plot service cut woman - difference
pdf("figures/Figure2_Panel3.pdf", width=4, height=7.1)
p = ggplot(service_cut_woman_diff, aes(y=point, x=treatments))
#p = p + coord_flip()
p = p + geom_hline(yintercept = 0,size=1 ,colour="grey50", linetype="dashed")
p = p + scale_y_continuous(name="Effect on share supporting greater service spending", limits=c(-.2, .2)
p = p + geom_pointrange(aes(ymin=lowerBH,ymax=upperBH), size= 1.5)

p = p + scale_x_discrete(name=paste("p-value of difference = ",service_cut_woman_pval, sep=""))
p = p + ggtitle("Effect of researcher gender on\nwillingness to support\npublic service expansion")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
############### End of code for Figures 1 and 2. ######################################

#### pull in counts of MTurk postings and generate Table 2
turk <- read.csv("MTurk_postings.csv")
head(turk)
```

```
##              Search.Term Accounts.Using.Real.Names Total.Accounts
## 1               Politics                         3              5
## 2                 Survey                       121            169
## 3               Research                        21             41
## 4               Academic                         5              9
## 5             University                         1              4
## 6     Political Science                          1              1
##   Proportion.Using.Name
## 1                  0.60
## 2                  0.72
## 3                  0.51
## 4                  0.56
## 5                  0.25
## 6                  1.00
```

```
print(xtable(turk), include.rownames=F)
```

```
## % latex table generated in R 3.2.3 by xtable 1.8-2 package
## % Fri Oct 20 15:23:21 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrr}
##   \hline
## Search.Term & Accounts.Using.Real.Names & Total.Accounts & Proportion.Using.Name \\
##   \hline
```

```
##     Politics  &   3 &   5 & 0.60 \\
##       Survey  & 121 & 169 & 0.72 \\
##       Research  &  21 &  41 & 0.51 \\
##       Academic  &   5 &   9 & 0.56 \\
##       University  &   1 &   4 & 0.25 \\
##       Political Science  &   1 &   1 & 1.00 \\
##       Psychology  &  23 &  31 & 0.74 \\
##       Economics  &   1 &   1 & 1.00 \\
##       TOTAL  & 176 & 261 & 0.67 \\
##    \hline
## \end{tabular}
## \end{table}
```

```
##########################################
### Supplementary information
##########################################

#################################
#### Data summaries
#################################

### Outcome Summaries
## List of variables
## (Table A3 in SI section A)
all.variables <- c("racial.resentment", "president.black.yes", "president.woman.yes","womenrole.equal",

outcome.summaries <- data.frame(Variable = c("Racial Resentment", "Would vote for a black president", "
                                Mean=sapply(all.variables, function(x) mean(researcher.data[[x]], na.rm=
                                Std.Dev = sapply(all.variables, function(x) sd(researcher.data[[x]], na
                                N.Responses = sapply(all.variables, function(x) sum(!is.na(researcher.da

print(xtable(outcome.summaries, caption = "Summary statistics for outcome variables", digits = 3), inclu
```

```
## % latex table generated in R 3.2.3 by xtable 1.8-2 package
## % Fri Oct 20 15:23:21 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrr}
##   \hline
## Variable & Mean & Std.Dev & N.Responses \\
##   \hline
## Racial Resentment & 0.468 & 0.280 & 1999 \\
##   Would vote for a black president & 0.934 & 0.249 & 2002 \\
##   Would vote for a woman president & 0.938 & 0.241 & 2003 \\
##   Women's role: Equal & 0.904 & 0.295 & 2004 \\
##   Women's role: Home & 0.053 & 0.225 & 2004 \\
##   Expand public services & 0.619 & 0.486 & 2005 \\
##   Cut public services & 0.381 & 0.486 & 2005 \\
##    \hline
## \end{tabular}
## \caption{Summary statistics for outcome variables}
## \end{table}
```

```
#### Covariates by treatment assignment
balance.tests <- data.frame(Variable = c("Prop. Male", "Prop. Democrat", "Prop. Republican", "Prop. Vot
```

```
                            N.Respondents = c(sum(!is.na(researcher.data$gender.male)), sum(!is.na(resea
                                              sum(!is.na(researcher.data$voteyes)), sum(!is.na(researche
                            Researcher.Male = NA, Researcher.Female = NA, t.stat.gender = NA, Researcher

balance.tests[1, c(3,4)] <- t.test(gender.male ~ treatment.woman, data=researcher.data)$estimate
balance.tests[1, c(5)] <- t.test(gender.male ~ treatment.woman, data=researcher.data)$statistic

balance.tests[1, c(6,7)] <- t.test(gender.male ~ treatment.black, data=researcher.data)$estimate
balance.tests[1, c(8)] <- t.test(gender.male ~ treatment.black, data=researcher.data)$statistic

balance.tests[2, c(3,4)] <- t.test(democrat ~ treatment.woman, data=researcher.data)$estimate
balance.tests[2, c(5)] <- t.test(democrat ~ treatment.woman, data=researcher.data)$statistic

balance.tests[2, c(6,7)] <- t.test(democrat ~ treatment.black, data=researcher.data)$estimate
balance.tests[2, c(8)] <- t.test(democrat ~ treatment.black, data=researcher.data)$statistic

balance.tests[3, c(3,4)] <- t.test(republican ~ treatment.woman, data=researcher.data)$estimate
balance.tests[3, c(5)] <- t.test(republican ~ treatment.woman, data=researcher.data)$statistic

balance.tests[3, c(6,7)] <- t.test(republican ~ treatment.black, data=researcher.data)$estimate
balance.tests[3, c(8)] <- t.test(republican ~ treatment.black, data=researcher.data)$statistic

balance.tests[4, c(3,4)] <- t.test(voteyes ~ treatment.woman, data=researcher.data)$estimate
balance.tests[4, c(5)] <- t.test(voteyes ~ treatment.woman, data=researcher.data)$statistic

balance.tests[4, c(6,7)] <- t.test(voteyes ~ treatment.black, data=researcher.data)$estimate
balance.tests[4, c(8)] <- t.test(voteyes ~ treatment.black, data=researcher.data)$statistic

balance.tests[5, c(3,4)] <- t.test(white ~ treatment.woman, data=researcher.data)$estimate
balance.tests[5, c(5)] <- t.test(white ~ treatment.woman, data=researcher.data)$statistic

balance.tests[5, c(6,7)] <- t.test(white ~ treatment.black, data=researcher.data)$estimate
balance.tests[5, c(8)] <- t.test(white ~ treatment.black, data=researcher.data)$statistic

balance.tests[6, c(3,4)] <- t.test(black ~ treatment.woman, data=researcher.data)$estimate
balance.tests[6, c(5)] <- t.test(black ~ treatment.woman, data=researcher.data)$statistic

balance.tests[6, c(6,7)] <- t.test(black ~ treatment.black, data=researcher.data)$estimate
balance.tests[6, c(8)] <- t.test(black ~ treatment.black, data=researcher.data)$statistic

balance.tests.gender <- balance.tests[,c(1,2,3,4,5)]
balance.tests.race <- balance.tests[,c(1,2,6,7,8)]

print(xtable(balance.tests.gender, caption = "Covariate means by treatment conditions - Researcher gende
```

```
## % latex table generated in R 3.2.3 by xtable 1.8-2 package
## % Fri Oct 20 15:23:21 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
##   \hline
## Variable & N.Respondents & Researcher.Male & Researcher.Female & t.stat.gender \\
##   \hline
## Prop. Male & 2005 & 0.508 & 0.468 & 1.782 \\
```

```
##   Prop. Democrat & 2004 & 0.437 & 0.456 & -0.839 \\
##   Prop. Republican & 2004 & 0.220 & 0.198 & 1.232 \\
##   Prop. Voted in 2012 & 2005 & 0.703 & 0.729 & -1.261 \\
##   Prop. White & 2002 & 0.820 & 0.805 & 0.837 \\
##   Prop. Black & 2002 & 0.070 & 0.065 & 0.389 \\
##    \hline
## \end{tabular}
## \caption{Covariate means by treatment conditions - Researcher gender}
## \end{table}
```

```
print(xtable(balance.tests.race, caption = "Covariate means by treatment conditions - Researcher race",
```

```
## % latex table generated in R 3.2.3 by xtable 1.8-2 package
## % Fri Oct 20 15:23:21 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
##   \hline
## Variable & N.Respondents & Researcher.White & Researcher.Black & t.stat.race \\
##   \hline
## Prop. Male & 2005 & 0.484 & 0.491 & -0.337 \\
##   Prop. Democrat & 2004 & 0.441 & 0.453 & -0.544 \\
##   Prop. Republican & 2004 & 0.206 & 0.211 & -0.307 \\
##   Prop. Voted in 2012 & 2005 & 0.714 & 0.718 & -0.191 \\
##   Prop. White & 2002 & 0.809 & 0.815 & -0.365 \\
##   Prop. Black & 2002 & 0.074 & 0.061 & 1.170 \\
##    \hline
## \end{tabular}
## \caption{Covariate means by treatment conditions - Researcher race}
## \end{table}
```

```
##############################################
### Completion rates (As in Section C of SI)
##############################################

#### Fit a regression model for treatment on completion
finish <- lm(finishedAndMatched ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + tre
### Bootstrap the vcov
set.seed(390)
boot.iter <- 5000
boot.estims <- matrix(ncol=length(coef(finish)), nrow=boot.iter)
for(it in 1:boot.iter){
  ### Resample
  researcher.data.boot <- researcher.data.full[sample(1:nrow(researcher.data.full), nrow(researcher.data
  finish.boot <- lm(finishedAndMatched ~ treatment.whitewoman + treatment.blackwoman + treatment.blackma
  boot.estims[it,] <- coef(finish.boot)
}
### Save the variance covariance matrix
finish_vcov <- cov(boot.estims)[-5,-5]

##### Visualize the results
prediction.frame.finish <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(0,1,0

finish_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(finish)$treatment.wh
                                            paste("Black Woman\nn=",sum(model.frame(finish)$treatment.bl
```

```
                                        paste("Black Man\nn=",sum(model.frame(finish)$treatment.bla
                                        paste("White Man\nn=",sum(model.frame(finish)$treatment.whi
                            point = predict(finish, newdata=prediction.frame.finish),
                            se = sqrt(c(finish_vcov[1,1] + finish_vcov[2,2] + 2*finish_vcov[1,2],
                                        finish_vcov[1,1] + finish_vcov[3,3] + 2*finish_vcov[1,3],
                                        finish_vcov[1,1] + finish_vcov[4,4] + 2*finish_vcov[1,4],
                                        finish_vcov[1,1])))
```

```
## Warning in predict.lm(finish, newdata = prediction.frame.finish):
## prediction from a rank-deficient fit may be misleading
```

```
finish_results$lower95 <- finish_results$point - qnorm(.975)*finish_results$se
finish_results$upper95 <- finish_results$point + qnorm(.975)*finish_results$se

#### Plot completion rates
pdf("figures/FigureA1.pdf", width=7, height=5)
p = ggplot(finish_results, aes(y=point, x=treatments)) ### Initialize plot
p = p + scale_y_continuous(name="% Completing Survey", limits=c(.96,1)) ### Y-axis
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 2) ###
p = p + scale_x_discrete(name=paste("Treatment conditions", sep=""))
p = p + ggtitle("Effect of treatments on survey non-completion rates")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
################################################
##### SI Section D
################################################

################################################
#### Do we have imbalance in the treatment conditions?
################################################

### Convert names to treatment labels
blackwomen <- c('Ebony Gaines', 'Deja Washington')

whitemen <- c('Brett Walsh', 'Connor Schroeder')

blackmen <- c('DeShawn Booker', 'Tyrone Robinson')

whitewomen <- c('Laurie Yoder', 'Molly Ryan')

researcher.data$treatVec <- NA
researcher.data$treatVec[researcher.data$treatment.whiteman == 1] <- "White Male"
researcher.data$treatVec[researcher.data$treatment.whitewoman == 1] <- "White Woman"
researcher.data$treatVec[researcher.data$treatment.blackman == 1] <- "Black Male"
researcher.data$treatVec[researcher.data$treatment.blackwoman == 1] <- "Black Woman"

set.seed(02138)
test.stat <- max(table(researcher.data$treatVec)) - min(table(researcher.data$treatVec))

total_num <- length(researcher.data$treatVec)
```

```
iter <- 100000
null_dist <- rep(NA, iter)

sim_dist <- rmultinom(iter, total_num, prob=c(.25, .25, .25, .25))
null_dist <- apply(sim_dist, 2, max) - apply(sim_dist, 2, min)

### p-value of hypothesis test for non-randomization. Our observed distribution is consistent with
### uniform randomization.
p.val <- mean(as.integer(null_dist > test.stat))


################ Are there differences in demographics across our treatment conditions?

### Gender
woman_select <- lm(gender.female ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + t
woman_select_vcov <- vcovHC(woman_select)
coeftest(woman_select, woman_select_vcov)
```

```
##
## t test of coefficients:
##
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.508046   0.024025 21.1463   <2e-16 ***
## treatment.whitewoman  0.013155   0.031932  0.4120   0.6804
## treatment.blackwoman  0.035244   0.033416  1.0547   0.2917
## treatment.blackman   -0.032031   0.032235 -0.9937   0.3205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(

woman_select_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(woman_select)$
                                                   paste("Black Woman\nn=",sum(model.frame(woman_select)$
                                                   paste("Black Man\nn=",sum(model.frame(woman_select)$t
                                                   paste("White Man\nn=",sum(model.frame(woman_select)$t
                                   point = predict(woman_select, newdata=prediction.frame.exploratory),
                                   se = sqrt(c(woman_select_vcov[1,1] + woman_select_vcov[2,2] + 2*woma
                                               woman_select_vcov[1,1] + woman_select_vcov[3,3] + 2*woma
                                               woman_select_vcov[1,1] + woman_select_vcov[4,4] + 2*woma
                                               woman_select_vcov[1,1])))
```

```
## Warning in predict.lm(woman_select, newdata =
## prediction.frame.exploratory): prediction from a rank-deficient fit may be
## misleading
```

```
woman_select_results$lower95 <- woman_select_results$point - qnorm(.975)*woman_select_results$se
woman_select_results$upper95 <- woman_select_results$point + qnorm(.975)*woman_select_results$se

woman_select_fstat <- 1- pf(summary(woman_select)[[10]][1], summary(woman_select)[[10]][2], summary(woma


#### Plot % women in-sample
pdf("figures/FigureA2.pdf", width=7, height=4)
p = ggplot(woman_select_results, aes(y=point, x=treatments))
p = p + geom_hline(yintercept = 0.5,size=1 ,colour="black", linetype="dotted")
```

```
p = p + scale_y_continuous(name="Share of female\nrespondents in sample", limits=c(.4, .6))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)
p = p + scale_x_discrete(name=paste("Treatment conditions\nF-test p-value = ", round(woman_select_fstat
p = p + ggtitle("Effect of researcher race and gender on\nselection into survey by gender")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
##################
# Make Figure A3 #
##################

# Panel 2: (Race - White)

white_select <- lm(white ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + treatment
white_select_vcov <- vcovHC(white_select)
coeftest(white_select, white_select_vcov)
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.8045977  0.0190550 42.2250   <2e-16 ***
## treatment.whitewoman   0.0077917  0.0251755  0.3095   0.7570
## treatment.blackwoman -0.0085023  0.0267721 -0.3176   0.7508
## treatment.blackman     0.0271953  0.0249534  1.0898   0.2759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(

white_select_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(white_select)$
                                                   paste("Black Woman\nn=",sum(model.frame(white_select)$
                                                   paste("Black Man\nn=",sum(model.frame(white_select)$t
                                                   paste("White Man\nn=",sum(model.frame(white_select)$t
                                   point = predict(white_select, newdata=prediction.frame.exploratory),
                                   se = sqrt(c(white_select_vcov[1,1] + white_select_vcov[2,2] + 2*white
                                               white_select_vcov[1,1] + white_select_vcov[3,3] + 2*white
                                               white_select_vcov[1,1] + white_select_vcov[4,4] + 2*white
                                               white_select_vcov[1,1])))
```

```
## Warning in predict.lm(white_select, newdata =
## prediction.frame.exploratory): prediction from a rank-deficient fit may be
## misleading
```

```
white_select_results$lower95 <- white_select_results$point - qnorm(.975)*white_select_results$se
white_select_results$upper95 <- white_select_results$point + qnorm(.975)*white_select_results$se


white_select_fstat <- 1- pf(summary(white_select)[[10]][1], summary(white_select)[[10]][2], summary(whi

#### Plot % white in-sample
pdf("figures/FigureA3_Panel2.pdf", width=7, height=4)
```

```
p = ggplot(white_select_results, aes(y=point, x=treatments))

p = p + scale_y_continuous(name="Share of white\nrespondents in sample", limits=c(.7, .9))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("Treatment conditions\nF-test p-value = ", round(white_select_fstat
p = p + ggtitle("Effect of researcher race and gender on\nselection into survey by race")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
# Panel 1: Race - Black

black_select <- lm(black ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + treatment
black_select_vcov <- vcovHC(black_select)
coeftest(black_select, black_select_vcov)
```

```
##
## t test of coefficients:
##
##                         Estimate Std. Error t value  Pr(>|t|)
## (Intercept)             0.087356   0.013569  6.4379 1.513e-10 ***
## treatment.whitewoman -0.023640   0.017032 -1.3880   0.16530
## treatment.blackwoman -0.020111   0.017910 -1.1229   0.26162
## treatment.blackman   -0.031903   0.016772 -1.9022   0.05729 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(

black_select_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(black_select)$
                                                   paste("Black Woman\nn=",sum(model.frame(black_select)$
                                                   paste("Black Man\nn=",sum(model.frame(black_select)$t
                                                   paste("White Man\nn=",sum(model.frame(black_select)$t
                                   point = predict(black_select, newdata=prediction.frame.exploratory),
                                   se = sqrt(c(black_select_vcov[1,1] + black_select_vcov[2,2] + 2*black
                                               black_select_vcov[1,1] + black_select_vcov[3,3] + 2*black
                                               black_select_vcov[1,1] + black_select_vcov[4,4] + 2*black
                                               black_select_vcov[1,1])))
```

```
## Warning in predict.lm(black_select, newdata =
## prediction.frame.exploratory): prediction from a rank-deficient fit may be
## misleading
```

```
black_select_results$lower95 <- black_select_results$point - qnorm(.975)*black_select_results$se
black_select_results$upper95 <- black_select_results$point + qnorm(.975)*black_select_results$se

black_select_fstat <- 1- pf(summary(black_select)[[10]][1], summary(black_select)[[10]][2], summary(blac

### Plot %black
pdf("figures/FigureA3_Panel1.pdf", width=7, height=4)
p = ggplot(black_select_results, aes(y=point, x=treatments))
p = p + scale_y_continuous(name="Share of black\nrespondents in sample", limits=c(0, .15))
```

```
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)
p = p + scale_x_discrete(name=paste("Treatment conditions\nF-test p-value = ", round(black_select_fstat
p = p + ggtitle("Effect of researcher race and gender on\nselection into survey by race")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
##################
# Make Figure A4 #
##################

# Panel 1: Democrats

dem_select <- lm(democrat ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + treatmen
dem_select_vcov <- vcovHC(dem_select)
coeftest(dem_select, dem_select_vcov)
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.4459770  0.0238877 18.6697   <2e-16 ***
## treatment.whitewoman -0.0095813  0.0317287 -0.3020   0.7627
## treatment.blackwoman  0.0334156  0.0333820  1.0010   0.3169
## treatment.blackman   -0.0160877  0.0320075 -0.5026   0.6153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(

dem_select_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(dem_select)$trea
                                                paste("Black Woman\nn=",sum(model.frame(dem_select)$trea
                                                paste("Black Man\nn=",sum(model.frame(dem_select)$treatm
                                                paste("White Man\nn=",sum(model.frame(dem_select)$treatm
                                 point = predict(dem_select, newdata=prediction.frame.exploratory),
                                 se = sqrt(c(dem_select_vcov[1,1] + dem_select_vcov[2,2] + 2*dem_select_
                                             dem_select_vcov[1,1] + dem_select_vcov[3,3] + 2*dem_select_
                                             dem_select_vcov[1,1] + dem_select_vcov[4,4] + 2*dem_select_
                                             dem_select_vcov[1,1])))
```

```
## Warning in predict.lm(dem_select, newdata = prediction.frame.exploratory):
## prediction from a rank-deficient fit may be misleading
```

```
dem_select_results$lower95 <- dem_select_results$point - qnorm(.975)*dem_select_results$se
dem_select_results$upper95 <- dem_select_results$point + qnorm(.975)*dem_select_results$se

dem_select_fstat <- 1- pf(summary(dem_select)[[10]][1], summary(dem_select)[[10]][2], summary(dem_selec

#### Plot % dem in-sample
pdf("figures/FigureA4_Panel1.pdf", width=7, height=4)
p = ggplot(dem_select_results, aes(y=point, x=treatments))

p = p + scale_y_continuous(name="Share of Democrat\nrespondents in sample", limits=c(.3, .6))
```

```
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("Treatment conditions\nF-test p-value = ", round(dem_select_fstat,4)
p = p + ggtitle("Effect of researcher race and gender on\nselection into survey by party ID")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
# Panel 2: Republicans

rep_select <- lm(republican ~ treatment.whitewoman + treatment.blackwoman + treatment.blackman + treatme
rep_select_vcov <- vcovHC(rep_select)
coeftest(rep_select, rep_select_vcov)
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.2068966  0.0194669 10.6281   <2e-16 ***
## treatment.whitewoman -0.0019496  0.0258431 -0.0754   0.9399
## treatment.blackwoman -0.0181764  0.0266931 -0.6809   0.4960
## treatment.blackman     0.0237308  0.0265998  0.8921   0.3724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(0

rep_select_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(rep_select)$trea
                                                paste("Black Woman\nn=",sum(model.frame(rep_select)$trea
                                                paste("Black Man\nn=",sum(model.frame(rep_select)$treatm
                                                paste("White Man\nn=",sum(model.frame(rep_select)$treatm
                                 point = predict(rep_select, newdata=prediction.frame.exploratory),
                                 se = sqrt(c(rep_select_vcov[1,1] + rep_select_vcov[2,2] + 2*rep_select_
                                             rep_select_vcov[1,1] + rep_select_vcov[3,3] + 2*rep_select_
                                             rep_select_vcov[1,1] + rep_select_vcov[4,4] + 2*rep_select_
                                             rep_select_vcov[1,1])))
```

```
## Warning in predict.lm(rep_select, newdata = prediction.frame.exploratory):
## prediction from a rank-deficient fit may be misleading
```

```
rep_select_results$lower95 <- rep_select_results$point - qnorm(.975)*rep_select_results$se
rep_select_results$upper95 <- rep_select_results$point + qnorm(.975)*rep_select_results$se

rep_select_fstat <- 1- pf(summary(rep_select)[[10]][1], summary(rep_select)[[10]][2], summary(rep_select

#### Plot % rep in-sample
pdf("figures/FigureA4_Panel2.pdf", width=7, height=4)
p = ggplot(rep_select_results, aes(y=point, x=treatments))

p = p + scale_y_continuous(name="Share of Republican\nrespondents in sample", limits=c(0.1, .3))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("Treatment conditions\nF-test p-value = ", round(rep_select_fstat,4)
```

```
p = p + ggtitle("Effect of researcher race and gender on\nselection into survey by party ID")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
##########################################
## Look at Attention Checks (SI section E)
##########################################

ac_pass <- lm(ac.pass.both ~ treatment.whiteman, data=researcher.data)

ac_pass_vcov <- vcovHC(ac_pass)
coeftest(ac_pass, ac_pass_vcov)
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error  t value Pr(>|t|)
## (Intercept)         0.8789809  0.0082365 106.7176   <2e-16 ***
## treatment.whiteman -0.0284062  0.0190096  -1.4943   0.1353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ac_pass_pval <- round(coeftest(ac_pass, ac_pass_vcov)[2,4], 3)
prediction.frame.ac <- data.frame(treatment.whiteman = c(1,0))

ac_pass_results <- data.frame(treatments = c(paste("White, Male Name\nn=",sum(model.frame(ac_pass)$trea
                                             paste("Other Name\nn=",sum(model.frame(ac_pass)$treatment.
                              point = predict(ac_pass, newdata=prediction.frame.ac),
                              se = sqrt(c(ac_pass_vcov[1,1] + ac_pass_vcov[2,2] + 2*ac_pass_vcov[1,2], a

ac_pass_results$lower95 <- ac_pass_results$point - qnorm(.975)*ac_pass_results$se
ac_pass_results$upper95 <- ac_pass_results$point + qnorm(.975)*ac_pass_results$se


#### Plot attention check completion (Figure A5 in SI)
pdf("figures/FigureA5.pdf", width=5, height=7)
p = ggplot(ac_pass_results, aes(y=point, x=treatments))

p = p + scale_y_continuous(name="% that successfully completed both attention checks", limits=c(.7, 1))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)

p = p + scale_x_discrete(name=paste("Treatment conditions\np-value of difference = ",ac_pass_pval, sep=
p = p + ggtitle("Effect of researcher gender\n and race on attention\n check completion")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```

```
#################################################################
#### Power calculations - using women equality as a hypothetical
```

```
#### (Reported in Section F of the SI)
#################################################################

#### Power calculations with multiple testing
### Parameters
set.seed(02138)
M <- 3
effect_size <- seq(from=0, to=.1, by=.0025)
sample_size <- 1000

iter <- 50000
alpha <- .05

### Variances
variances <- c(var(researcher.data.womenequal$womenrole.equal),
               var(researcher.data.womanpres$president.woman.yes),
               var(researcher.data.servicesmore$services.more) )*1.5 ### The third one is greater than
### Strict max at .25
variances[3] <- .25

##########################
### Store power calc result
power_results <- rep(0, length(effect_size))
names(power_results) <- effect_size

##### All effects true
### For each effect_size
for (size in effect_size){
  reject <- rep(0, iter)
  #### Draw test statistics
  for (i in 1:iter){
    ts_variance <- variances/sample_size + variances/sample_size
    effects <- rep(size, M)
    test_stats <- rnorm(3, effects/sqrt(ts_variance), 1)

    ## Sample true test statistic
    all_pvals <- pnorm(-abs(test_stats), 0, 1)*2

    ## Order pvalues
    ordered_stats <- sort(all_pvals)

    ## Which ordered statistic is the ``true" one

    ## Do we reject?
    thresholds <- ((1:M)/M)*alpha

    ## If nothing is significant
    if (sum(ordered_stats < thresholds) == 0){
      reject[i] <- 0
    }else{

      max_thresh <- max((1:M)[ordered_stats < thresholds])
      reject[i] <- max_thresh/M
```

```
    }

  }
  power_results[as.character(size)] <- mean(reject)
}

power_results_all <- power_results

### Variances
variances <- c(var(researcher.data.womenequal$womenrole.equal),
               var(researcher.data.womanpres$president.woman.yes),
               var(researcher.data.servicesmore$services.more) )*1.5

### Strict max at .25
variances[3] <- .25

############## One effect is true, others are 0

set.seed(02138)
##########################
### Store power calc result
power_results <- rep(0, length(effect_size))
names(power_results) <- effect_size

##### One effect is true
### For each effect_size
for (size in effect_size){
  reject <- rep(0, iter)
  #### Draw test statistics
  for (i in 1:iter){
    ts_variance <- variances/sample_size + variances/sample_size
    effects <- c(0, 0, size)
    test_stats <- rnorm(3, effects/sqrt(ts_variance), 1)


    ## Sample true test statistic
    all_pvals <- pnorm(-abs(test_stats), 0, 1)*2
    key_pval = all_pvals[3]

    ## Order pvalues
    ordered_stats <- sort(all_pvals)

    ## Which ordered statistic is the ``true" one
    key_element = (1:M)[which(ordered_stats == key_pval)]

    ## Do we reject?
    thresholds <- ((1:M)/M)*alpha

    ## If nothing is significant
    if (sum(ordered_stats < thresholds) == 0){
      reject[i] <- 0
    }else{
```

```
      max_thresh <- max((1:M)[ordered_stats < thresholds])
      if (key_element <= max_thresh){
        reject[i] <- 1
      }else{
        reject[i] <- 0
      }
    }

  }
  power_results[as.character(size)] <- mean(reject)
}

power_results_1 <- power_results

################################################

### Plot curves: this is Figure A6 in SI

pdf("figures/FigureA6.pdf", width=7, height=5)
plot(y=power_results_all, x=seq(from=0, to=.1, by=.0025), type="l", lwd=3,
     ylab = expression(paste("Power (", alpha, "= .05), two-sided, 3 tests, BH corrected")), xlab = "Abs
lines(y=power_results_1, x=seq(from=0, to=.1, by=.0025), type="l", lwd=3, lty=2)
abline(h=0.8, lty=3, lwd=2, col="darkgrey")
abline(h=0.5, lty=3, lwd=2, col="darkgrey")
abline(h=1, lty=3, lwd=2, col="darkgrey")

legend("bottomright", lty=c(1,2), legend=c("All Nulls False", "One Null False"), lwd=3)
dev.off()
```

```
## pdf
##   2
```

```
#####################################################################
#### Exploratory components (As in Section G of SI)
#####################################################################

################
### Finer-grained treatment
################

researcher.data$womenrole[researcher.data$womenrole == -99] <- NA
researcher.data$womenrole.equal <- ifelse(researcher.data$womenrole == 1, 1, 0)
researcher.data$womenrole.home <- ifelse(researcher.data$womenrole == 2, 1, 0)
researcher.data$womenrole.dk <- ifelse(researcher.data$womenrole == 3, 1, 0)

exploratory_women_equal <- lm(womenrole.equal ~ treatment.whitewoman + treatment.blackwoman + treatment
exploratory_women_equal_vcov <- vcovHC(exploratory_women_equal)
coeftest(exploratory_women_equal, exploratory_women_equal_vcov)
```

```
##
## t test of coefficients:
##
##                       Estimate Std. Error t value  Pr(>|t|)
## (Intercept)           0.877880   0.015753 55.7272 < 2.2e-16 ***
## treatment.whitewoman 0.051448   0.019095  2.6944  0.007111 **
```

```
## treatment.blackwoman 0.022553   0.021049  1.0714  0.284101
## treatment.blackman   0.022489   0.020354  1.1049  0.269340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
exploratory_women_equal_pval <- round(coeftest(exploratory_women_equal, exploratory_women_equal_vcov)[2


prediction.frame.exploratory <- data.frame(treatment.whitewoman = c(1,0,0,0), treatment.blackwoman = c(

exploratory_women_equal_results <- data.frame(treatments = c(paste("White Woman\nn=",sum(model.frame(exp
                                            paste("Black Woman\nn=",sum(model.frame(exploratory_women_e
                                            paste("Black Man\nn=",sum(model.frame(exploratory_women_equa
                                            paste("White Man\nn=",sum(model.frame(exploratory_women_equa
                                point = predict(exploratory_women_equal, newdata=prediction.frame.explorat
                                se = sqrt(c(exploratory_women_equal_vcov[1,1] + exploratory_women_equal_vc
                                            exploratory_women_equal_vcov[1,1] + exploratory_women_equal_vc
                                            exploratory_women_equal_vcov[1,1] + exploratory_women_equal_vc
                                            exploratory_women_equal_vcov[1,1])))
```

```
## Warning in predict.lm(exploratory_women_equal, newdata =
## prediction.frame.exploratory): prediction from a rank-deficient fit may be
## misleading
```

```
exploratory_women_equal_results$lower95 <- exploratory_women_equal_results$point - qnorm(.975)*explorat
exploratory_women_equal_results$upper95 <- exploratory_women_equal_results$point + qnorm(.975)*explorat

#### Plot exploratory - white woman vs. white man on equality
pdf("figures/FigureA7.pdf", width=7, height=4)
p = ggplot(exploratory_women_equal_results, aes(y=point, x=treatments))
p = p + scale_y_continuous(name="% that say women should\nbe equal to men", limits=c(.84, .96))
p = p + geom_pointrange(aes(ymin=lower95,ymax=upper95), size= 1.5)
p = p + scale_x_discrete(name=paste("Treatment conditions\np-value of white woman vs. white man differe
p = p + ggtitle("Effect of researcher race and gender on\nattitudes on gender equality")
p = p + theme_bw1()
print(p)
dev.off()
```

```
## pdf
##   2
```