Reproduction Code for Supplementary Tables S1-S10

Analysis for 'Symbolic Incentives in High-Stakes Prosocial Behavior'

June 18, 2025

Contents

Introduction

This document provides the R code to reproduce Supplementary Tables S1 through S10. The code reads from `reproducible_dataset.csv`, a cleaned dataset containing only the variables necessary for this analysis.

```
# Load the clean, reproducible dataset
df <- read_csv('reproducible_dataset.csv')

# Define the two main analysis samples for convenience
baseline_sample <- df
study_sample <- df %>% filter(is_study_sample == TRUE)
```

Table S1: Study Sample Representativeness

This table compares the characteristics of the study sample to the general US population.

```
# Create a summary of the study sample
s1_study_summary <- study_sample %>%
  summarise(
    `Female` = mean(is_female, na.rm = TRUE),
    `White` = mean(race == "White", na.rm = TRUE),
    `Hispanic` = mean(race == "Hispanic", na.rm = TRUE),
    `Other Race` = mean(race == "Other", na.rm = TRUE),
    `Black` = mean(race == "Black", na.rm = TRUE),
    `Asian` = mean(race == "Asian", na.rm = TRUE),
    `Age (years)` = mean(age, na.rm = TRUE),
    `Prior Plasma Donation` = mean(prior_plasma_donation, na.rm = TRUE),
    `AB+ or AB- Blood Type` = mean(is_ab_blood_type, na.rm = TRUE),
    `Annual Donation Avg.` = mean(avg_annual_donations, na.rm = TRUE),
    `Median Income ($)` = mean(zip_median_income, na.rm = TRUE),
    `Midwest Region` = mean(us_region == "Midwest", na.rm = TRUE),
    `Northeast Region` = mean(us_region == "Northeast", na.rm = TRUE),
    `South Region` = mean(us_region == "South", na.rm = TRUE),
    `West Region` = mean(us_region == "West", na.rm = TRUE),
    `Urban Region` = mean(is_urban, na.rm = TRUE),
    `COVID-19 Incidence` = mean(county_covid_per_100k, na.rm = TRUE)
  ) %>%
  pivot_longer(everything(), names_to = "Characteristic", values_to = "Study Sample")

# Manually create the US Population data frame
us_pop_data <- tribble(
  ~Characteristic,             ~`US Population`,
  "Female",                    0.508, "White",                     0.763,
  "Hispanic",                  0.185, "Other Race",                0.03,
  "Black",                     0.134, "Asian",                     0.06,
  "Age (years)",               38.10, "Prior Plasma Donation",     NA,
  "AB+ or AB- Blood Type",     0.040, "Annual Donation Avg.",      NA,
  "Median Income ($)",         67521.00, "Midwest Region",            0.21,
  "Northeast Region",          0.17, "South Region",              0.38,
  "West Region",               0.24, "Urban Region",              0.807,
  "COVID-19 Incidence",        4224
)

# Join the two summaries and display the table using kable
table_s1 <- left_join(s1_study_summary, us_pop_data, by = "Characteristic")
kable(table_s1, caption = "Table S1: Study Sample Representativeness", digits = 3, format = "latex", bo
```

Table 1: Table S1: Study Sample Representativeness

Characteristic	Study Sample	US Population
Female	0.500	0.508
White	0.873	0.763
Hispanic	0.062	0.185
Other Race	0.022	0.030
Black	0.021	0.134
Asian	0.019	0.060
Age (years)	46.477	38.100
Prior Plasma Donation	0.030	NA
AB+ or AB- Blood Type	0.045	0.040
Annual Donation Avg.	1.727	NA
Median Income ($)	74978.586	67521.000
Midwest Region	0.407	0.210
Northeast Region	0.159	0.170
South Region	0.281	0.380
West Region	0.153	0.240
Urban Region	0.904	0.807
COVID-19 Incidence	4426.087	4224.000

Table S2: Balance Check

This table checks for balance on covariates between the treatment and control groups.

```
# Panel A: Study Sample Balance
panel_a <- study_sample %>%
  group_by(treatment_group) %>%
  summarise(
    Female = mean(is_female), White = mean(race == "White"), Age = mean(age),
    `Prior Plasma Donation` = mean(prior_plasma_donation),
    `Annual Donation Avg.` = mean(avg_annual_donations),
    `Median Income` = mean(zip_median_income), `Urban Region` = mean(is_urban),
    `COVID-19 Incidence` = mean(county_covid_per_100k)
  ) %>%
  pivot_longer(-treatment_group, names_to = "Characteristic") %>%
  pivot_wider(names_from = treatment_group, values_from = value)

kable(panel_a, caption = "Table S2 (Panel A): Study Sample Demographics Balance", digits = 2, format =
```

```
# Panel B: Baseline Sample Selection Criteria Balance
panel_b <- baseline_sample %>%
  group_by(treatment_group) %>%
  summarise(`Opened Email` = mean(email_opened), `No Phone Contact` = mean(no_phone_contact)) %>%
  pivot_longer(-treatment_group, names_to = "Characteristic") %>%
  pivot_wider(names_from = treatment_group, values_from = value)

kable(panel_b, caption = "Table S2 (Panel B): Sample Selection Criteria Balance", digits = 2, format =
```

Table 2: Table S2 (Panel A): Study Sample Demographics Balance

Characteristic	Control	Symbolic Incentives
Female	0.49	0.51
White	0.87	0.88
Age	45.71	47.25
Prior Plasma Donation	0.04	0.02
Annual Donation Avg.	1.74	1.72
Median Income	NA	NA
Urban Region	NA	NA
COVID-19 Incidence	NA	4481.73

Table 3: Table S2 (Panel B): Sample Selection Criteria Balance

Characteristic	Control	Symbolic Incentives
Opened Email	0.45	0.45
No Phone Contact	0.23	0.23

Table S3: Attrition Information

This table shows the number of donors enrolled, analyzed, and excluded.

```
table_s3 <- baseline_sample %>%
  mutate(ExclusionReason = case_when(
    is_study_sample == FALSE & email_opened == 0 & no_phone_contact == 0 ~ "Excluded (Both)",
    is_study_sample == FALSE & email_opened == 0 ~ "Excluded (Did Not Open)",
    is_study_sample == FALSE & no_phone_contact == 0 ~ "Excluded (Phone Contact)",
    TRUE ~ "Analyzed (Study Sample)"
  )) %>%
  group_by(treatment_group) %>%
  count(ExclusionReason) %>%
  pivot_wider(names_from = ExclusionReason, values_from = n)

kable(table_s3, caption = "Table S3: Attrition Information", format = "latex", booktabs = TRUE)
```

Table 4: Table S3: Attrition Information

treatment_group	Analyzed (Study Sample)	Excluded (Both)	Excluded (Did Not Open)	Excluded (Phone Conta
Control	1165	4606	1235	3
Symbolic Incentives	1157	4549	1206	3

Table S4: Main Analysis (Average Marginal Effects)

This table shows the Average Marginal Effects (AMEs) from logistic regressions on the primary outcomes.

```
# Define the control variables for the models
controls <- "is_female + age + race + prior_plasma_donation + bloodtype_group + avg_annual_donations + 

# Run models
m1 <- glm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave)")), data = study_sa
m2 <- glm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave) +", controls)), data
m3 <- glm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave)")), data = study_sample
m4 <- glm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave) +", controls)), data = 

# Helper function to extract and format AMEs correctly
get_ame <- function(model) {
  s <- summary(margins(model))
  # Find the treatment effect row
  treatment_row <- s %>% filter(grepl("treatment_group", factor))

  # Format the estimate with stars and the standard error
  stars <- case_when(treatment_row$p < 0.01 ~ "*", treatment_row$p < 0.05 ~ "*", treatment_row$p < 0.1 
  estimate_str <- paste0(sprintf("%.4f", treatment_row$AME), stars)
  se_str <- paste0("(", sprintf("%.4f", treatment_row$SE), ")")

  return(c(estimate_str, se_str))
}

# Manually build the results table to ensure correctness
results_table <- data.frame(
  Term = c("Symbolic Incentives", ""),
  `(1) Appt. Creation` = get_ame(m1),
  `(2) Appt. Creation` = get_ame(m2),
  `(3) Donation` = get_ame(m3),
  `(4) Donation` = get_ame(m4)
)

# Add GOF stats
gof_rows <- data.frame(
  Term = c("Num.Obs.", "Wave FE", "Controls"),
  `(1) Appt. Creation` = c(nobs(m1), "Yes", "No"),
  `(2) Appt. Creation` = c(nobs(m2), "Yes", "Yes"),
  `(3) Donation` = c(nobs(m3), "Yes", "No"),
  `(4) Donation` = c(nobs(m4), "Yes", "Yes")
)

# Combine and print using kable
final_table_s4 <- bind_rows(results_table, gof_rows)
kable(final_table_s4, caption = "Table S4: Effect of Symbolic Incentives (AMEs from Logistic Regression
```

Table 5: Table S4: Effect of Symbolic Incentives (AMEs from Logistic Regression)

Term	X.1..Appt..Creation	X.2..Appt..Creation	X.3..Donation	X.4..Donation
Symbolic Incentives	0.0147+	0.0193*	0.0018	0.0016
	(0.0080)	(0.0081)	(0.0021)	(0.0023)
Num.Obs.	2322	2086	2322	2086
Wave FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Table 6: Table S5: ITT Analysis on Baseline Population (AMEs)

Term	X.1..Appt..Creation	X.2..Appt..Creation	X.3..Donation	X.4..Donation
Symbolic Incentives	0.0026+	0.0033*	0.0009	0.0015
	(0.0016)	(0.0016)	(0.0011)	(0.0011)
Num.Obs.	20916	18771	20916	18771
Wave FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Table S5: Intent-to-Treat (ITT) Analysis

This table repeats the main analysis on the full baseline population.

```
# Run models on the baseline sample
m1_itt <- glm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave)")), data = basel
m2_itt <- glm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave) +", controls)),
m3_itt <- glm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave)")), data = baseline_
m4_itt <- glm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave) +", controls)), data

# Manually build the results table
results_table_itt <- data.frame(
  Term = c("Symbolic Incentives", ""),
  `(1) Appt. Creation` = get_ame(m1_itt),
  `(2) Appt. Creation` = get_ame(m2_itt),
  `(3) Donation` = get_ame(m3_itt),
  `(4) Donation` = get_ame(m4_itt)
)

gof_rows_itt <- data.frame(
  Term = c("Num.Obs.", "Wave FE", "Controls"),
  `(1) Appt. Creation` = c(nobs(m1_itt), "Yes", "No"),
  `(2) Appt. Creation` = c(nobs(m2_itt), "Yes", "Yes"),
  `(3) Donation` = c(nobs(m3_itt), "Yes", "No"),
  `(4) Donation` = c(nobs(m4_itt), "Yes", "Yes")
)

final_table_s5 <- bind_rows(results_table_itt, gof_rows_itt)
kable(final_table_s5, caption = "Table S5: ITT Analysis on Baseline Population (AMEs)", booktabs = TRUE
```

Table 7: Table S6: Effect of Symbolic Incentives (OLS with Robust SEs)

	(1) Appt. Creation	(2) Appt. Creation	(3) Donation	(4) Donation
Symbolic Incentives	0.0147+	0.0176*	0.0018	0.0013
	(0.0080)	(0.0083)	(0.0021)	(0.0020)
Num.Obs.	2322	2086	2322	2086
Wave FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Table S6: OLS Regressions

This table shows the main results using OLS with robust standard errors.

```
# Run OLS models
m1_ols <- lm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave)")), data = study
m2_ols <- lm(as.formula(paste("appointment_within_48h ~ treatment_group + factor(wave) +", controls)), 
m3_ols <- lm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave)")), data = study_sam
m4_ols <- lm(as.formula(paste("donated_within_13d ~ treatment_group + factor(wave) +", controls)), data

models_ols <- list(
  "(1) Appt. Creation" = m1_ols, "(2) Appt. Creation" = m2_ols,
  "(3) Donation" = m3_ols, "(4) Donation" = m4_ols
)

modelsummary(models_ols,
             title = "Table S6: Effect of Symbolic Incentives (OLS with Robust SEs)",
             fmt = 4, estimate = "{estimate}{stars}", statistic = "({std.error})",
             vcov = "HC1",
             coef_map = c("treatment_groupSymbolic Incentives" = "Symbolic Incentives"),
             gof_map = c("nobs"),
             add_rows = data.frame(
               "term" = c("Wave FE", "Controls"),
               "(1) Appt. Creation" = c("Yes", "No"), "(2) Appt. Creation" = c("Yes", "Yes"),
               "(3) Donation" = c("Yes", "No"), "(4) Donation" = c("Yes", "Yes")
             ),
             output = "latex")
```

Table 8: Table S7: Effect of Symbolic Incentives over Different Time Windows (Logit Coefficients)

	(1) Appt. (24h)	(2) Appt. (7d)	(3) Appt. (Anytime)	(4) Donation (Anytime)
Symbolic Incentives	0.5633*	0.3820+	0.1385	0.0538
	(0.2626)	(0.2160)	(0.1681)	(0.1964)
Num.Obs.	2086	2086	2086	2086
Wave FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Table S7: Different Time Windows

This table shows logistic regression coefficients for outcomes measured over different time horizons.

```
# Run models with different DVs
m1_tf <- glm(as.formula(paste("appointment_within_24h ~ treatment_group + factor(wave) +", controls)), c
m2_tf <- glm(as.formula(paste("appointment_within_7d ~ treatment_group + factor(wave) +", controls)), da
m3_tf <- glm(as.formula(paste("appointment_anytime ~ treatment_group + factor(wave) +", controls)), data
m4_tf <- glm(as.formula(paste("donated_anytime ~ treatment_group + factor(wave) +", controls)), data = s

models_tf <- list(
  "(1) Appt. (24h)" = m1_tf, "(2) Appt. (7d)" = m2_tf,
  "(3) Appt. (Anytime)" = m3_tf, "(4) Donation (Anytime)" = m4_tf
)

modelsummary(models_tf,
             title = "Table S7: Effect of Symbolic Incentives over Different Time Windows (Logit Coeffi
             fmt = 4, estimate = "{estimate}{stars}", statistic = "({std.error})",
             coef_map = c("treatment_groupSymbolic Incentives" = "Symbolic Incentives"),
             gof_map = c("nobs"),
             add_rows = data.frame(
               "term" = c("Wave FE", "Controls"),
               "(1) Appt. (24h)" = c("Yes", "Yes"), "(2) Appt. (7d)" = c("Yes", "Yes"),
               "(3) Appt. (Anytime)" = c("Yes", "Yes"), "(4) Donation (Anytime)" = c("Yes", "Yes")
             ),
             output = "latex")
```

Table 9: Table S8: Effect on Unsubscription and Total Donations (OLS with Robust SEs)

	(1) Unsubscribe	(2) Unsubscribe	(3) Total Donations	(4) Total Donations
Symbolic Incentives	−0.0017	−0.0017	0.0229	0.0083
	(0.0021)	(0.0017)	(0.0295)	(0.0300)
Num.Obs.	2322	2086	2322	2086
Wave FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Table S8: Unsubscription and Total Donations

This table shows OLS results for secondary outcomes.

```
# Run models for secondary outcomes
m1_sec <- lm(as.formula(paste("unsubscribe ~ treatment_group + factor(wave)")), data = study_sample)
m2_sec <- lm(as.formula(paste("unsubscribe ~ treatment_group + factor(wave) +", controls)), data = study
m3_sec <- lm(as.formula(paste("total_donations_post_treatment ~ treatment_group + factor(wave)")), data
m4_sec <- lm(as.formula(paste("total_donations_post_treatment ~ treatment_group + factor(wave) +", cont

models_sec <- list(
  "(1) Unsubscribe" = m1_sec, "(2) Unsubscribe" = m2_sec,
  "(3) Total Donations" = m3_sec, "(4) Total Donations" = m4_sec
)

modelsummary(models_sec,
             title = "Table S8: Effect on Unsubscription and Total Donations (OLS with Robust SEs)",
             fmt = 4, estimate = "{estimate}{stars}", statistic = "({std.error})",
             vcov = "HC1",
             coef_map = c("treatment_groupSymbolic Incentives" = "Symbolic Incentives"),
             gof_map = c("nobs"),
             add_rows = data.frame(
               "term" = c("Wave FE", "Controls"),
               "(1) Unsubscribe" = c("Yes", "No"), "(2) Unsubscribe" = c("Yes", "Yes"),
               "(3) Total Donations" = c("Yes", "No"), "(4) Total Donations" = c("Yes", "Yes")
             ),
             output = "latex")
```

Table S9: Heterogeneous Treatment Effects

This table explores whether the treatment effect varies across pre-registered subgroups.

```
# Define models with interaction terms
hte_models <- list(
  "Female" = lm(appointment_within_48h ~ treatment_group * is_female + factor(wave), data = study_sampl
  "High Prior Donations" = lm(appointment_within_48h ~ treatment_group * high_prior_donations + factor(
  "Urban" = lm(appointment_within_48h ~ treatment_group * is_urban + factor(wave), data = study_sample)
  "Majority Dem. County" = lm(appointment_within_48h ~ treatment_group * is_majority_dem_county + facto
)

modelsummary(hte_models,
             title = "Table S9: Heterogeneous Treatment Effects on Appointment Creation",
             fmt = 3, estimate = "{estimate}{stars}", statistic = "({std.error})",
             vcov = "HC1",
             coef_rename = c(
               "treatment_groupSymbolic Incentives" = "Treatment Effect (Symbolic Incentives)",
               "is_femaleTRUE" = "Baseline (Female)",
               "high_prior_donationsTRUE" = "Baseline (High Donor)",
               "is_urbanTRUE" = "Baseline (Urban)",
               "is_majority_dem_countyTRUE" = "Baseline (Dem. County)",
               "treatment_groupSymbolic Incentives:is_femaleTRUE" = "Interaction: Trt x Female",
               "treatment_groupSymbolic Incentives:high_prior_donationsTRUE" = "Interaction: Trt x High
               "treatment_groupSymbolic Incentives:is_urbanTRUE" = "Interaction: Trt x Urban",
               "treatment_groupSymbolic Incentives:is_majority_dem_countyTRUE" = "Interaction: Trt x De
             ),
             gof_map = "nobs", output = "latex")
```

Table 10: Table S9: Heterogeneous Treatment Effects on Appointment Creation

	Female	High Prior Donations	Urban	Majority I
(Intercept)	0.004	0.010	−0.016	−0
	(0.009)	(0.008)	(0.013)	(0
Treatment Effect (Symbolic Incentives)	0.006	0.003	−0.002	0.0
	(0.011)	(0.011)	(0.014)	(0
is_female	−0.001			
	(0.010)			
factor(wave)2	0.026**	0.024**	0.022*	0.0
	(0.009)	(0.009)	(0.010)	(0
factor(wave)3	0.039**	0.038**	0.032*	0.0
	(0.012)	(0.012)	(0.013)	(0
factor(wave)4	0.039***	0.039***	0.038**	0.0
	(0.011)	(0.011)	(0.012)	(0
factor(wave)5	−0.008	−0.010	−0.011	−0
	(0.007)	(0.007)	(0.008)	(0
treatment_groupSymbolic Incentives:is_female	0.018			
	(0.016)			
Baseline (High Donor)		−0.012		
		(0.010)		
Interaction: Trt x High Donor		0.024		
		(0.016)		
is_urban			0.022+	
			(0.012)	
treatment_groupSymbolic Incentives:is_urban			0.018	
			(0.017)	
is_majority_dem_county				0.0
				(0
treatment_groupSymbolic Incentives:is_majority_dem_county				−0
				(0
Num.Obs.	2322	2322	2091	2

Table 11: Table S10: Effect on Intention-Behavior Gap (Logit Coefficients)

	(1) No Controls	(2) With Controls
Symbolic Incentives	0.274	0.353
	(0.496)	(0.518)
Num.Obs.	91	91
Wave FE	Yes	Yes
Controls	No	Yes

Table S10: Intention-Behavior Gap

This table analyzes the effect of the treatment on the gap between scheduling an appointment and actually donating.

```
# Filter for the relevant population: those who made an appointment
gap_sample <- study_sample %>% filter(!is.na(intention_behavior_gap))

# Run models
m1_gap <- glm(intention_behavior_gap ~ treatment_group + factor(wave), data = gap_sample, family = "bin
m2_gap <- glm(intention_behavior_gap ~ treatment_group + factor(wave) + is_female + age + race, data = g

models_gap <- list(
  "(1) No Controls" = m1_gap,
  "(2) With Controls" = m2_gap
)

modelsummary(models_gap,
             title = "Table S10: Effect on Intention-Behavior Gap (Logit Coefficients)",
             fmt = 3, estimate = "{estimate}{stars}", statistic = "({std.error})",
             coef_map = c("treatment_groupSymbolic Incentives" = "Symbolic Incentives"),
             gof_map = c("nobs"),
             add_rows = data.frame(
               "term" = c("Wave FE", "Controls"),
               "(1) No Controls" = c("Yes", "No"),
               "(2) With Controls" = c("Yes", "Yes")
             ),
             output = "latex")
```