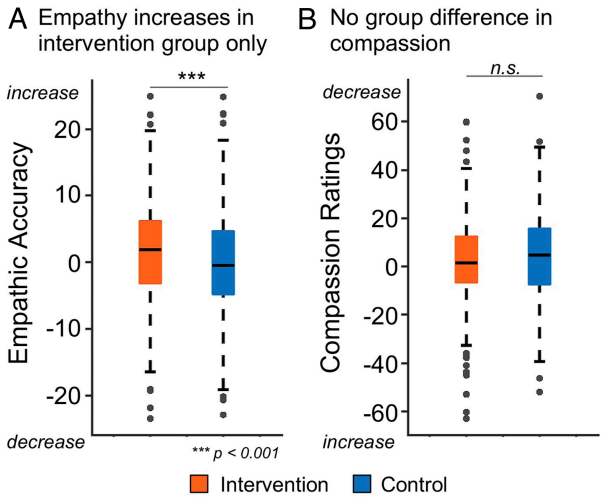
figure_assignment

Sihan Yang

Original Paper

- link
  - paper: https://www.pnas.org/doi/10.1073/pnas.2322819121
  - data and code: https://osf.io/eugjd/
- introduction
- main figure

```
knitr::include_graphics("paper/fig01.jpg")
```



  - explanation
  - strengths
    - has provided the essential information about the main hypothesis
  - weakness

* not showing the distribution.
* not color-blindness or black-and-white printing friendly
* confusing labels
* redundant title
* the overall balance of size of different components and the layout is not pretty
* the zero line could be further stressed so that it will be easier to tell whether it's increasing or decreasing

Reproduction of Main Figures

Dara Preprocessing

While the authors have provided the preprocessed data for further analysis in R, here we started with the raw data to examine whether their data are processed correctly. (Since the raw empathy inference response are not provided, we will use directly used the accuracy metrics (Pearson correlation score and RMSE) provided by the authors).

```
library(tidyverse)

data_path <- 'data/N709_EmpathicAccuracyTaskDat.csv'
loaded <- read.csv(data_path)
```

```
# get rid of redundant columns
cleared_loaded <- loaded |>
  select(obsID, movie, gender, age, obsRace, Ideology, SES,
         stimID, stimRace, storyteller_label_attn_check, cond, visit, EAcorr, EArmse,
         compassion, RemoveMisLabeled1)

# remove those with missing data
cleared_loaded <- cleared_loaded |>
  filter(!RemoveMisLabeled1)
cleared_loaded <- cleared_loaded |> select(-RemoveMisLabeled1)
```

Separate the subject's empathy data and other data

```r
subject_info <- cleared_loaded |>
  select(obsID, movie, gender, age, obsRace, Ideology, SES) |>
  distinct(obsID, .keep_all = TRUE)

survey_data <- cleared_loaded |>
  select(obsID, stimID, storyteller_label_attn_check, cond, visit, EAcorr, EArmse, compassion

head(subject_info)
```

```
  obsID      movie gender age obsRace     Ideology SES
1   271  Moneyball  woman  29   White      Liberal   8
2   274 Concussion    man  50   White      Liberal NaN
3   276  Moneyball    man  65   White   OtherRight   6
4   284 Just Mercy  woman  49   White      Liberal NaN
5   286 Concussion  woman  64   White      Liberal   2
6   289 Just Mercy    man  56   Asian Conservative   3
```

Compute each subjects' average inference accuracy and compassion in two visits

```r
empathy_collapsed <- survey_data |>
  drop_na() |>
  group_by(obsID, visit, storyteller_label_attn_check) |>
  summarize (
    compassion = mean(compassion, na.rm=TRUE),
    EAcorr=mean(EAcorr, na.rm=TRUE),
    EArmse=mean(EArmse, na.rm=TRUE)
  ) |> ungroup()
```

```
`summarise()` has grouped output by 'obsID', 'visit'. You can override using
the `.groups` argument.
```

```r
# leave out those who do not have both types of story-teller in both visits
# i.e. visit 1/2 x story-telley prisonser/student
empathy_collapsed <- empathy_collapsed |>
  group_by(obsID) |>
  filter(n() == 4)  |>
  ungroup()

nrow(empathy_collapsed)
```

```
[1] 2668
```

```
head(empathy_collapsed)
```

```
# A tibble: 6 x 6
  obsID visit storyteller_label_attn_check compassion  EAcorr EArmse
  <int> <int> <chr>                             <dbl>   <dbl>  <dbl>
1   271     1 Formerly Incarcerated              86.8  0.626    17.7
2   271     1 Student                            94.2  0.354    25.2
3   271     2 Formerly Incarcerated              88.3  0.138    29.2
4   271     2 Student                            92.8  0.461    26.2
5   273     1 Formerly Incarcerated              56.7  0.645    17.0
6   273     1 Student                            43   -0.0404   29.2
```

Fitting

Examine how the interaction between time point (i.e. 'visit'), the condition (i.e. 'cond', what type movie people watch between two surveys) and label (i.e. whether the story teller is labeled as 'formerly incarcerated' or 'student') affect RMSE score (as in the original paper). Here we closely follow how the original study code categorical data.

```
library(lmerTest)
library(broom.mixed)

# combine all info, and rename some columns to prepare for fititng
fitting_table <- subject_info |>
  inner_join(empathy_collapsed, by='obsID') |>
  mutate(visit = case_when(
    visit == 1 ~ "pre",
    visit == 2 ~ "post",
  )) |>
  mutate(
    cond = if_else(movie == "Just Mercy", "intervention", "control")
  ) |>
  rename(storyteller_label=storyteller_label_attn_check)

# remove any with nan
fitting_table <- fitting_table |>
  drop_na()

# further clean up (e.g. some of more than one race --> more)
```

```r
possible_races <- c("White", "Asian", "Hispanic or Latino", "Black or African American", "Nat
fitting_table <- fitting_table |>
  mutate(
    obsRace = if_else(obsRace %in% possible_races, obsRace, "More")
  ) |>
  mutate(
    gender = if_else(gender == "nonbinary", "other", gender)
  )

# convert some columns to categories
cols_to_factorize <- c("obsID", "gender", "obsRace", "Ideology", "SES")
fitting_table <- fitting_table |>
  mutate(across(all_of(cols_to_factorize), as.factor)) |>
  mutate(cond=factor(cond, levels=c("control", "intervention"))) |>
  mutate(visit=factor(visit, levels=c("pre", "post"))) |>
  mutate(storyteller_label=factor(storyteller_label, levels=c("Student", "Formerly Incarcerat

# apply contrasts
contrasts(fitting_table$cond) = contr.poly(2)
contrasts(fitting_table$visit) = contr.poly(2)
contrasts(fitting_table$storyteller_label) = contr.poly(2)
contrasts(fitting_table$obsRace) = contr.poly(6)
contrasts(fitting_table$gender) = contr.poly(3)
contrasts(fitting_table$Ideology) = contr.poly(4)
contrasts(fitting_table$SES) = contr.poly(10)

# fit full lme model ()
rmse_fit_model <- lmer(
  EArmse ~ cond* storyteller_label * visit
  + (1|obsID) + obsRace + gender + Ideology + SES,
  data=fitting_table)
rmse_fit_result <- tidy(rmse_fit_model, effects = "fixed", conf.int = TRUE)
```

```r
rmse_fit_result |> mutate(across(where(is.double), ~round(., 4)))
```

```
# A tibble: 27 x 9
   effect term     estimate std.error statistic    df p.value conf.low conf.high
   <chr>  <chr>       <dbl>     <dbl>     <dbl> <dbl>   <dbl>    <dbl>     <dbl>
 1 fixed  (Interc~   27.6       1.02      27.0    580  0        25.6     29.7
 2 fixed  cond.L     -0.597     0.295     -2.03   580  0.0432   -1.18    -0.0182
 3 fixed  storyte~   -0.338     0.164     -2.06  1797  0.0392   -0.660   -0.0167
```

```
 4 fixed  visit.L    -0.515     0.164     -3.14  1797  0.0017   -0.837   -0.194
 5 fixed  obsRace~    3.42      1.94       1.76   580  0.0784   -0.389    7.22
 6 fixed  obsRace~   -1.50      0.863     -1.73   580  0.0837   -3.19     0.2
 7 fixed  obsRace~   -6.42      2.73      -2.35   580  0.019   -11.8     -1.06
 8 fixed  obsRace~   -7.79      2.96      -2.63   580  0.0087  -13.6     -1.98
 9 fixed  obsRace~   -4.39      1.75      -2.51   580  0.0124   -7.84    -0.953
10 fixed  gender.L   -0.628     0.304     -2.06   580  0.0397   -1.23    -0.0298
# i 17 more rows
```

Check the main interaction

- Label: storyteller label (s: student; p: former prisoner)
- Condition: whether subject was assigned to intervention (i: intervention; c: control)
- Time: whether the survey was done before or after watching a film (1: before; 2: after)

```
mapping <- c(
  "(Intercept)" = "Intercept",
  "visit.L" = "Time",
  "cond.L" = "Condition",
  "storyteller_label.L" = "Label",
  "cond.L:visit.L" = "Time*Condition",
  "cond.L:storyteller_label.L" = "Condition*Label",
  "storyteller_label.L:visit.L" = "Time*Label",
  "cond.L:storyteller_label.L:visit.L"="Time*Condition*Label"
)

# Filter and map terms
selected_result <- rmse_fit_result |>
  filter(term %in% names(mapping)) |>
  mutate(term = recode(term, !!!mapping)) |>
  mutate(across(where(is.double), ~ round(., 3))) # easier to check
selected_result
```

```
# A tibble: 8 x 9
  effect term      estimate std.error statistic    df p.value conf.low conf.high
  <chr>  <chr>        <dbl>     <dbl>     <dbl> <dbl>   <dbl>    <dbl>     <dbl>
1 fixed  Intercept   27.6       1.02     27.0     580   0       25.6      29.7
2 fixed  Condition   -0.597     0.295    -2.03    580   0.043   -1.18     -0.018
3 fixed  Label       -0.338     0.164    -2.06   1797   0.039   -0.66     -0.017
4 fixed  Time        -0.515     0.164    -3.14   1797   0.002   -0.837    -0.194
5 fixed  Conditio~   -0.24      0.232    -1.03   1797   0.302   -0.695     0.215
```

```
6 fixed  Time*Con~   -0.595     0.232    -2.57   1797   0.01    -1.05     -0.14
7 fixed  Time*Lab~   -0.1       0.232    -0.433  1797   0.665   -0.555     0.355
8 fixed  Time*Con~   -0.674     0.328    -2.05   1797   0.04    -1.32     -0.03
```

However...Also the original study does not adjust their p-value...

Visualization

First compute how people emotion inference accuracy and compassion changed after watching the film

```
# first compute how rating changes before and after watching a film
ea_change_table <- empathy_collapsed |>
  inner_join(subject_info |> select(obsID, movie), by="obsID") |>
  pivot_wider(
    names_from = visit,
    values_from = c(compassion, EArmse, EAcorr)
  ) |>
  mutate(
    compassion_diff = compassion_2 - compassion_1,
    corr_diff = EAcorr_2 - EAcorr_1,
    rmse_diff = EArmse_2 - EArmse_1
  ) |> mutate(
    acc_corr_diff = corr_diff,
    acc_rmse_diff = -rmse_diff
  ) |> mutate (
    cond = if_else(movie == "Just Mercy", "intervention", "control")
  ) |>
  select(obsID, cond, storyteller_label_attn_check, compassion_diff, acc_corr_diff, acc_rmse
```

T-test

A fast test of whether RMSE, correlation and compasion significantly increase or decrease, which is one of the important hypothesis to test.

```
test_increase_decrease <- function(data, col) {
  ttest <- t.test(data[[col]], mu=0);
  result <- tibble(
    mean = mean(data[[col]], na.rm = TRUE),
    t_stat = ttest$statistic,
    p_value = ttest$p.value,
```

```
    conf_low = ttest$conf.int[1],
    conf_high = ttest$conf.int[2]
  )
  result
}
```

- RMSE

```
# Test differences for each storyteller_label
rmse_ttests <- ea_change_table |>
  group_by(storyteller_label_attn_check, cond) |>
  summarise(
    test_results = list(test_increase_decrease(cur_data(), "acc_rmse_diff"))
  ) |>
  unnest(test_results) |>
  mutate(across(where(is.double), ~ round(., 3)))
rmse_ttests
```

```
# A tibble: 4 x 7
# Groups:   storyteller_label_attn_check [2]
  storyteller_label_attn_check cond       mean t_stat p_value conf_low conf_high
  <chr>                        <chr>     <dbl>  <dbl>   <dbl>    <dbl>     <dbl>
1 Formerly Incarcerated        control  -0.301  -0.73   0.466   -1.11      0.509
2 Formerly Incarcerated        interve~  1.81    4.17   0        0.957     2.66
3 Student                      control   0.506   1.12   0.265   -0.386     1.40
4 Student                      interve~  0.742   1.59   0.113   -0.177     1.66
```

The result suggests that emotion inference accuracy (measured by RMSE) change only significantly when story teller is labeled as 'Formerly Incarcerated' and the movie watched between the two surveys is the intervention one ('Just Mercy').

- CORR

```
# Test differences for each storyteller_label
corr_ttests <- ea_change_table |>
  group_by(storyteller_label_attn_check, cond) |>
  summarise(
    test_results = list(test_increase_decrease(cur_data(), "acc_corr_diff"))
  ) |>
  unnest(test_results) |>
  mutate(across(where(is.double), ~ round(., 3)))
```

```
print(corr_ttests)
```

```
# A tibble: 4 x 7
# Groups:   storyteller_label_attn_check [2]
  storyteller_label_attn_check cond       mean t_stat p_value conf_low conf_high
  <chr>                        <chr>     <dbl>  <dbl>   <dbl>    <dbl>     <dbl>
1 Formerly Incarcerated        control  -0.023 -1.29    0.197   -0.058     0.012
2 Formerly Incarcerated        interve~  0.029  1.56    0.121   -0.008     0.066
3 Student                      control  -0.003 -0.162   0.871   -0.041     0.035
4 Student                      interve~  0.027  1.38    0.17    -0.012     0.066
```

Interestingly, the effect disappeared if instead pearson correlation is used to measure emotion inference accuracy

- Compassion

```
# Test differences for each storyteller_label
compassion_ttests <- ea_change_table |>
  group_by(storyteller_label_attn_check, cond) |>
  summarise(
    test_results = list(test_increase_decrease(cur_data(), "compassion_diff"))
  ) |>
  unnest(test_results) |>
  mutate(across(where(is.double), ~ round(., 3)))
compassion_ttests
```

```
# A tibble: 4 x 7
# Groups:   storyteller_label_attn_check [2]
  storyteller_label_attn_check cond       mean t_stat p_value conf_low conf_high
  <chr>                        <chr>     <dbl>  <dbl>   <dbl>    <dbl>     <dbl>
1 Formerly Incarcerated        control   -4.36  -4.56   0        -6.24    -2.48
2 Formerly Incarcerated        interven~ -2.82  -2.76   0.006    -4.83    -0.813
3 Student                      control   -3.72  -4.03   0        -5.53    -1.90
4 Student                      interven~ -2.79  -2.34   0.02     -5.13    -0.442
```

Compassion overall decreases significantly in the second survey, regardless of number of the label of story-teller and the type of movies watched.

Finally, test whether there are group differences by two-sample ttest

```r
former_incarcerated <- ea_change_table |>
  filter(storyteller_label_attn_check == 'Formerly Incarcerated')

rmse_compare_ttest <- t.test(acc_rmse_diff ~ cond, data = former_incarcerated)
print(rmse_compare_ttest)
```

```
    Welch Two Sample t-test

data:  acc_rmse_diff by cond
t = -3.5284, df = 655.04, p-value = 0.0004473
alternative hypothesis: true difference in means between group control and group interventio
95 percent confidence interval:
 -3.2864196 -0.9363921
sample estimates:
     mean in group control mean in group intervention
                -0.3006644                  1.8107414
```

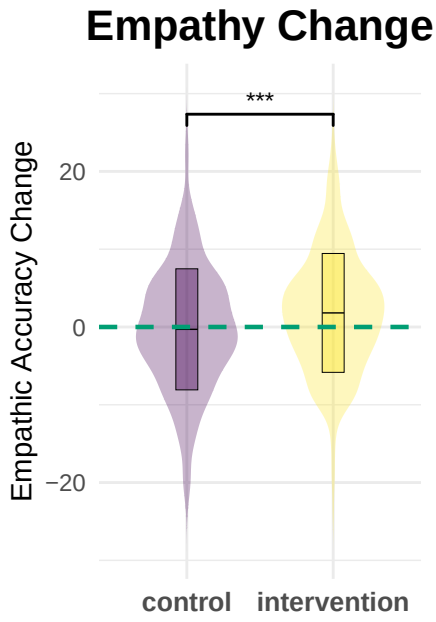
```r
compassion_compare_ttest <- t.test(compassion_diff ~ cond, data = former_incarcerated)
print(compassion_compare_ttest)
```

```
    Welch Two Sample t-test

data:  compassion_diff by cond
t = -1.1, df = 652.84, p-value = 0.2717
alternative hypothesis: true difference in means between group control and group interventio
95 percent confidence interval:
 -4.282368  1.207146
sample estimates:
     mean in group control mean in group intervention
                -4.358310                 -2.820699
```

This suggests intervention only brings a significant difference for empathy but not compassion for former prisoner.

Replicate figure 1

```r
library(ggplot2)
library(ggsignif)
library(viridis)


ggplot(former_incarcerated, aes(x = cond, y = acc_rmse_diff, fill = cond)) +
  geom_violin(trim = FALSE, alpha = 0.3, width=0.7, color=NA) +
  stat_summary(fun.data = mean_sdl, geom = "crossbar", , fun.args=list(mult=1), width = 0.15
  geom_signif(comparisons = list(c("control", "intervention")), map_signif_level = TRUE, text
  geom_hline(yintercept = 0, linetype = "dashed", color="#009E73", size = 0.8) +
  labs(title = "Empathy Change", x = "", y = "Empathic Accuracy Change") +
  scale_fill_viridis_d(option="viridis") +
  theme_minimal() +
  theme(
    legend.position="none",
    aspect.ratio=1.6,
    plot.title=element_text(size=16, face = "bold", hjust = 0.5),
    axis.text.x = element_text(size = 10, face = "bold")
  ) +
  theme(legend.position = "none")
```




```r
ggplot(former_incarcerated, aes(x = cond, y = compassion_diff, fill = cond)) +
  geom_violin(trim = FALSE, alpha = 0.3, width=0.7, color=NA) +
```

```r
  stat_summary(fun.data = mean_sdl, geom = "crossbar", , fun.args=list(mult=1), width = 0.15
  geom_signif(comparisons = list(c("control", "intervention")), map_signif_level = TRUE, text
  geom_hline(yintercept = 0, linetype = "dashed", color="#009E73", size = 0.8) +
  labs(title = "Compassion Change", x = "", y = "Compassion Rating Change") +
  scale_fill_viridis_d(option="viridis") +
  theme_minimal() +
  theme(
    legend.position="none",
    aspect.ratio=1.6,
    plot.title=element_text(size=16, face = "bold", hjust = 0.5),
    axis.text.x = element_text(size = 10, face = "bold")
  ) +
  theme(legend.position = "none")
```