Replication Analysis

David John Baker

24/06/2020

This document provides the analysis for "Lyrics' Emotional Influences in Songs: A Conceptual Replication of Ali and Peynircioğlu's (2006) Study". The script creates both figures for the paper as well as the statistical analysis.

- [ ] Add in Analyses about Inconsistencies in 2,3,4
- [ ] Re-order factor Levels in both data

```
# Libraries Needed
library(ggridges)
library(ez)
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
library(multcomp)
```

```
## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(sjPlot)
```

```
## #refugeeswelcome
```

```r
library(patchwork)
```

```
##
## Attaching package: 'patchwork'

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

```r
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(magrittr)
library(readr)
library(ggplot2)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':
##
##     extract

## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack
```

```r
library(gt)

# Paper Data

table_1_values <- c(7.85, 8.18, 6.62, 6.91,
                    5.79, 5.20, 6.51, 6.11,
                    7.75, 7.95, 4.05, 5.41,
                    5.00, 4.41, 6.08, 6.59)

grim_test <- function(sample_size, likert_start, likert_end, means){
  integer_values <- sample_size * means
  bottom <- round(floor(integer_values)/ sample_size, 2)
  top <- round(ceiling(integer_values)/ sample_size, 2)
  df <- data.frame(bottom, top, means)
  df$sample_size <- sample_size
  df$possible <- ifelse(df$bottom == means | df$top == means, yes = TRUE,no = FALSE)
  df
}

table_1 <- grim_test(32, 1, 9, table_1_values)


table_1$gender <- rep(c("woman","man"),8)
table_1$lyrics <- rep(c("no lyric", "no lyric", "lyric","lyric"),4)
table_1$emotions <- c(rep("happy",4), rep("sad",4), rep("calm",4), rep("angry",4))

table_1 <- table_1 %>%
  tibble() %>%
    mutate(valence = case_when(
    emotions == "happy" ~ "positive",
    emotions == "calm" ~ "positive",
    emotions == "angry" ~ "negative",
    emotions == "sad" ~ "negative"
  ))


# LSU Data
df_complete <- read_csv("Complete_Data.csv", col_types = cols(age = col_character()))

df_complete <- df_complete %>%
  mutate(age = str_replace_all(string = age, pattern = "eighteen", replacement = "18")) %>%
  mutate(age = str_replace_all(string = age, pattern = "3", replacement = NA_character_)) %>%
  mutate(age = as.numeric(age))

valence_table_exact <- read_csv("data/valence_sheet.csv")
```

```
## Parsed with column specification:
## cols(
##   title = col_character(),
##   Condition = col_character(),
##   Track_ID = col_double(),
##   Valence = col_character()
## )
```

```r
df_complete %>%
  left_join(valence_table_exact) %>%
  rename(track_valence = Valence) -> df_complete
```

```
## Joining, by = c("title", "Condition", "Track_ID")
```

```r
df_complete <- df_complete %>%
   mutate(global_valence = case_when(
    track_valence == "Happy" ~ "positive",
    track_valence == "Calm" ~ "positive",
    track_valence == "Angry" ~ "negative",
    track_valence == "Sad" ~ "negative"
  ))

gender_table <- df_complete %>%
  select(subject, gender) %>%
  distinct() %>%
  mutate(gender = str_to_lower(gender)) %>%
  mutate(cleaned_gender = case_when(
    gender == "cis male" ~ "male",
    gender == "false" ~ NA_character_,
    gender == "female" ~ "female",
    gender == "femalen" ~ "female",
    gender == "male" ~ "male",
    gender == "malen" ~ "male"
  )) %>%
  select(subject,cleaned_gender)
```

Before running analysis on our own data, below we visualize the results that were reported in the original paper so we know if we match them.

AP Data

In the A+P 2006 paper, under Experiment 1, the authors report results of4 x 2 x 2 ANOVA with Lyrics, Emotion, and Gender:

- [ ] Main Effect of Emotion with Happy sigdif from other three, other three no-sigdif between
- [ ] Main Effect of Valence (Positive Better than Negative) (Should have been reported as contrast)
- [ ] Main Effect of Lyrics (Lyrics Higher)
- [ ] NO Main effect of Gender
- [ ] Lyric and Gender Interaction (lyrics made motions of melody for women worse but not men)
- [ ] Sanity Check of Congruency Always Higher

```r
table_1$emotions <- factor(table_1$emotions, levels = c("happy","calm","angry","sad"))

# Main Effect of Intended Emotion
table_1 %>%
  group_by(emotions) %>%
  summarise(mean_ratings = mean(means))  %>%
  ggplot(aes(y = mean_ratings,
             x = emotions)) +
```

```r
  geom_bar(stat = "identity") +
  labs(title = "Main Effect of Intended Emotion",
       subtitle = "Happy Significant Diff, but Not Others") +
  theme_minimal() -> me_intended_emotion
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Main Effect of Valence
table_1 %>%
  group_by(valence) %>%
  summarise(mean_ratings = mean(means))  %>%
  ggplot(aes(y = mean_ratings,
             x = valence)) +
  geom_bar(stat = "identity") +
  labs(title = "Main Effect of Valence") +
  theme_minimal() -> me_valence
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Main Effect of Lyrics
table_1 %>%
  group_by(lyrics) %>%
  summarise(mean_ratings = mean(means))  %>%
  ggplot(aes(y = mean_ratings,
             x = lyrics)) +
  geom_bar(stat = "identity") +
  labs(title = "Main Effect of Lyrics") +
  theme_minimal() -> me_lyrics
```

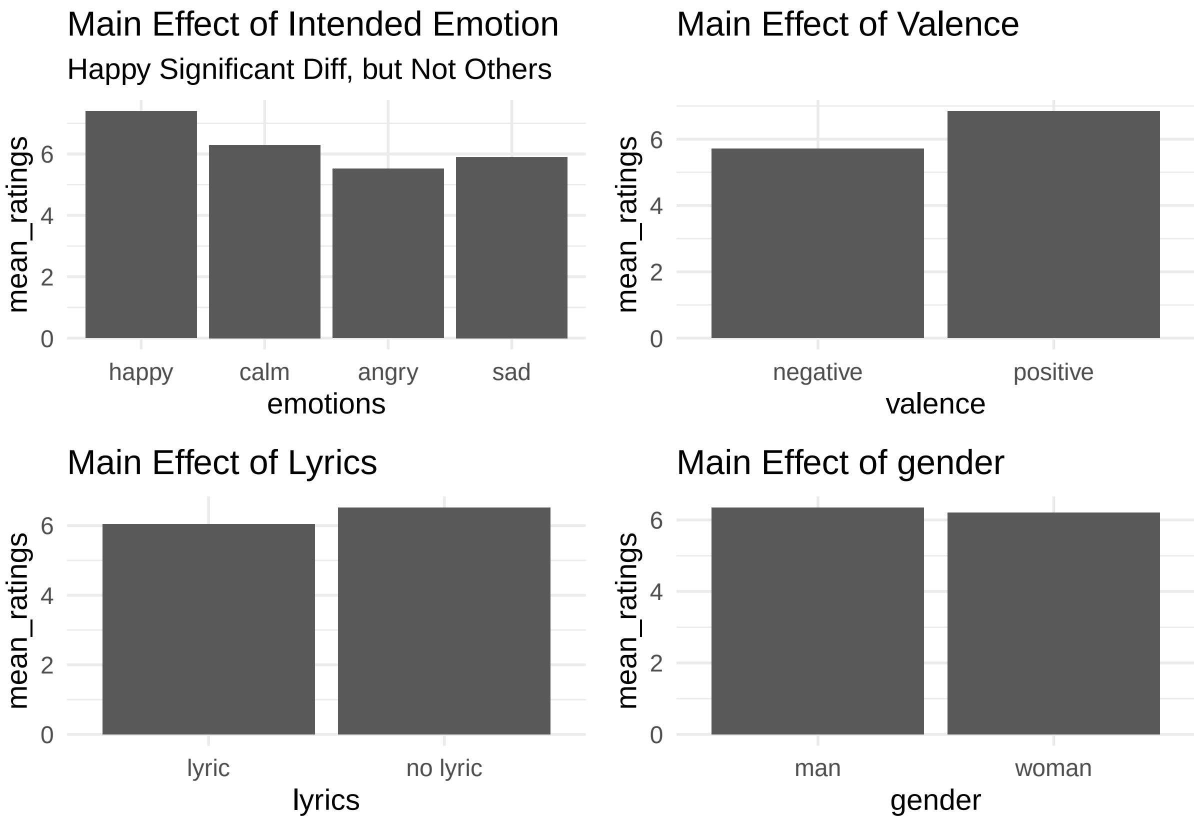
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# NO Main Effect of Gender

table_1 %>%
  group_by(gender) %>%
  summarise(mean_ratings = mean(means))  %>%
  ggplot(aes(y = mean_ratings,
             x = gender)) +
  geom_bar(stat = "identity") +
  labs(title = "Main Effect of gender") +
  theme_minimal() -> me_gender
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
(ap_panel <- ( me_intended_emotion | me_valence ) / ( me_lyrics | me_gender ))
```



LSU Data

First Visualise:

```
# Standard Deviation of Mean Function
std <- function(x) sd(x)/sqrt(length(x))

# Create Long Data, Subset Only Congruancy Conditions
df_congruent <- df_complete %>%
  select(subject, Happy:AngryStressful, Condition, track_valence, global_valence) %>%
  pivot_longer(cols = Happy:AngryStressful, names_to = "response",values_to = "rating") %>%
  # Congrency Add Here!!
  mutate(congru = track_valence == response) %>%
  filter(congru == TRUE)

df_congruent$response <- factor(df_congruent$response, levels = c("Happy","Calm","AngryStressful","Sad"

# Main Effect of Intended Emotion
me_intended_emotion_lsu <- df_congruent %>%
  group_by(response) %>%
  summarise(
    response_mean = mean(rating),
    response_sd = sd(rating),
    response_std = std(rating)
  ) %>%
```

```r
  ggplot(aes(x = response, y = response_mean)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
    geom_errorbar(aes(ymin = response_mean + response_std,
                      ymax = response_mean - response_std)) +
  labs(title = "Main Effect of Emotion")
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Main Effect of Valence
me_valence_lsu <- df_congruent %>%
  group_by(global_valence) %>%
  summarise(
    response_mean = mean(rating),
    response_sd = sd(rating),
    response_std = std(rating)
  ) %>%
  ggplot(aes(x = global_valence, y = response_mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = response_mean + response_std,
                    ymax = response_mean - response_std)) +
  theme_minimal() +
  labs(title = "Main Effect of Valence")
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Main Effect of Lyrics
me_lyrics_lsu <- df_congruent %>%
  group_by(Condition) %>%
  summarise(
    response_mean = mean(rating),
    response_sd = sd(rating),
    response_std = std(rating)
  ) %>%
  ggplot(aes(x = Condition, y = response_mean)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  geom_errorbar(aes(ymin = response_mean + response_std,
                    ymax = response_mean - response_std)) +
  labs(title = "Main Effect of Condition")
```

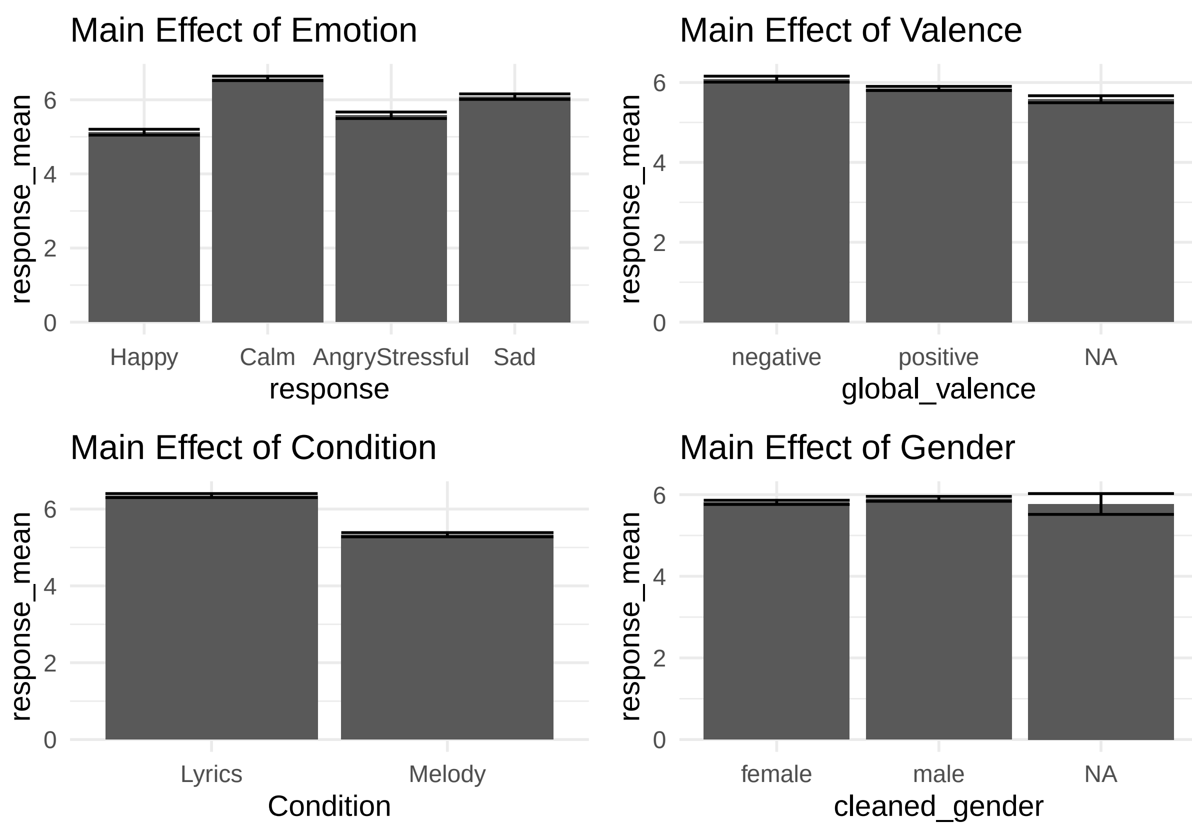
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# Main Effect of Gender
me_gender_lsu <- df_congruent %>%
  left_join(gender_table) %>%
  group_by(cleaned_gender) %>%
  summarise(
    response_mean = mean(rating),
    response_sd = sd(rating),
    response_std = std(rating)
  ) %>%
```

```r
  ggplot(aes(x = cleaned_gender, y = response_mean)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  geom_errorbar(aes(ymin = response_mean + response_std,
                    ymax = response_mean - response_std)) +
  labs(title = "Main Effect of Gender")
```

```
## Joining, by = "subject"
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
(lsu_panel <- (me_intended_emotion_lsu | me_valence_lsu) / (me_lyrics_lsu | me_gender_lsu))
```



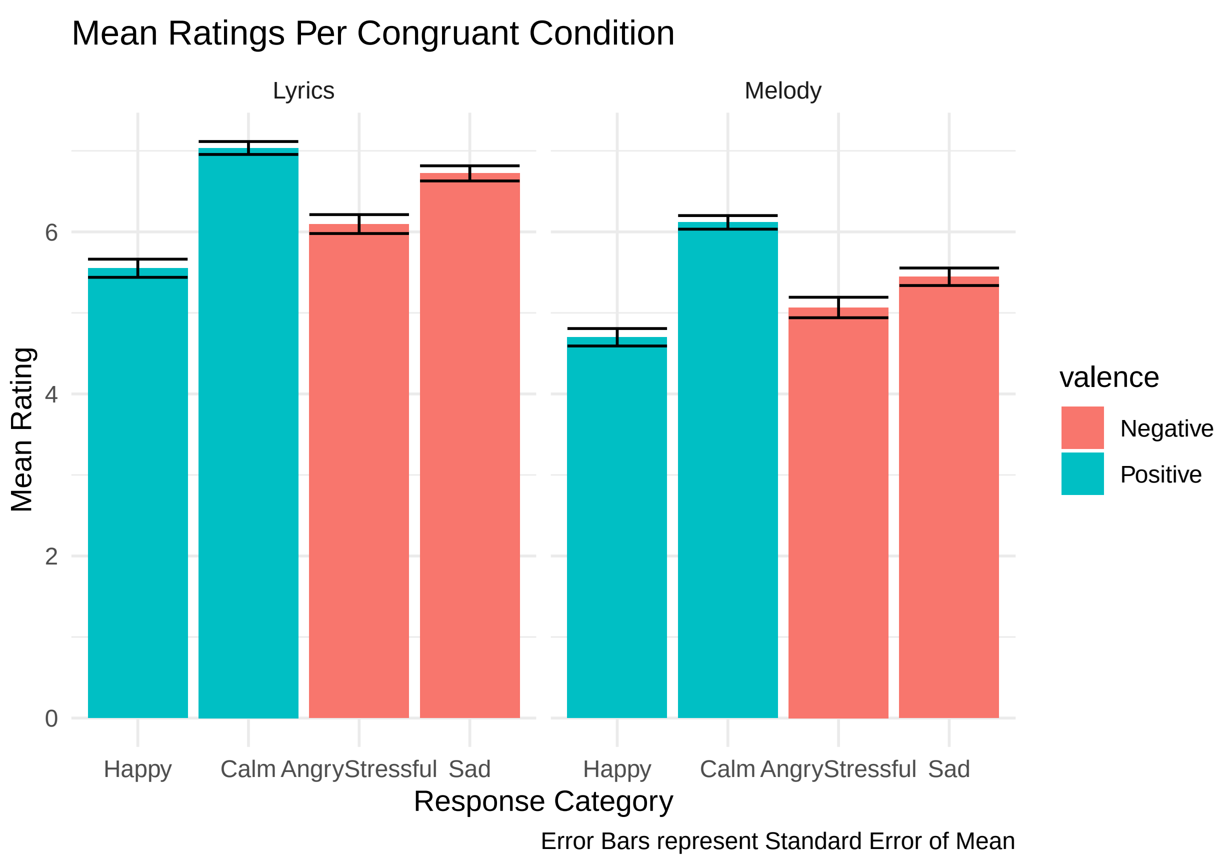

```r
# Big Plots of LSU
# Make Mean Rating Per Condition
df_congruent  %>%
  left_join(gender_table) %>%
  group_by(Condition, response) %>%
  summarise(mean_rating = mean(rating),
            sd_rating = sd(rating),
            max_rating = max(rating),
            min_rating = min(rating),
            se_rating = std(rating),
            number_of_ratings = n()) %>%
  mutate(response = factor(response, levels = c("Happy","Calm","AngryStressful","Sad"))) %>%
```

```r
mutate(valence = case_when(
  response == "AngryStressful" ~ "Negative",
  response == "Sad" ~ "Negative",
  response == "Happy" ~ "Positive",
  response == "Calm" ~ "Positive"

)) %>%
ggplot(aes(x = response, y = mean_rating, fill = valence)) +
geom_bar(stat = "identity") +
facet_wrap(~Condition) +
geom_errorbar(aes(ymin = mean_rating + se_rating, ymax = mean_rating - se_rating)) +
labs(title = "Mean Ratings Per Congruant Condition",
     caption = "Error Bars represent Standard Error of Mean",
     x = "Response Category",
     y = "Mean Rating") +
theme_minimal()
```

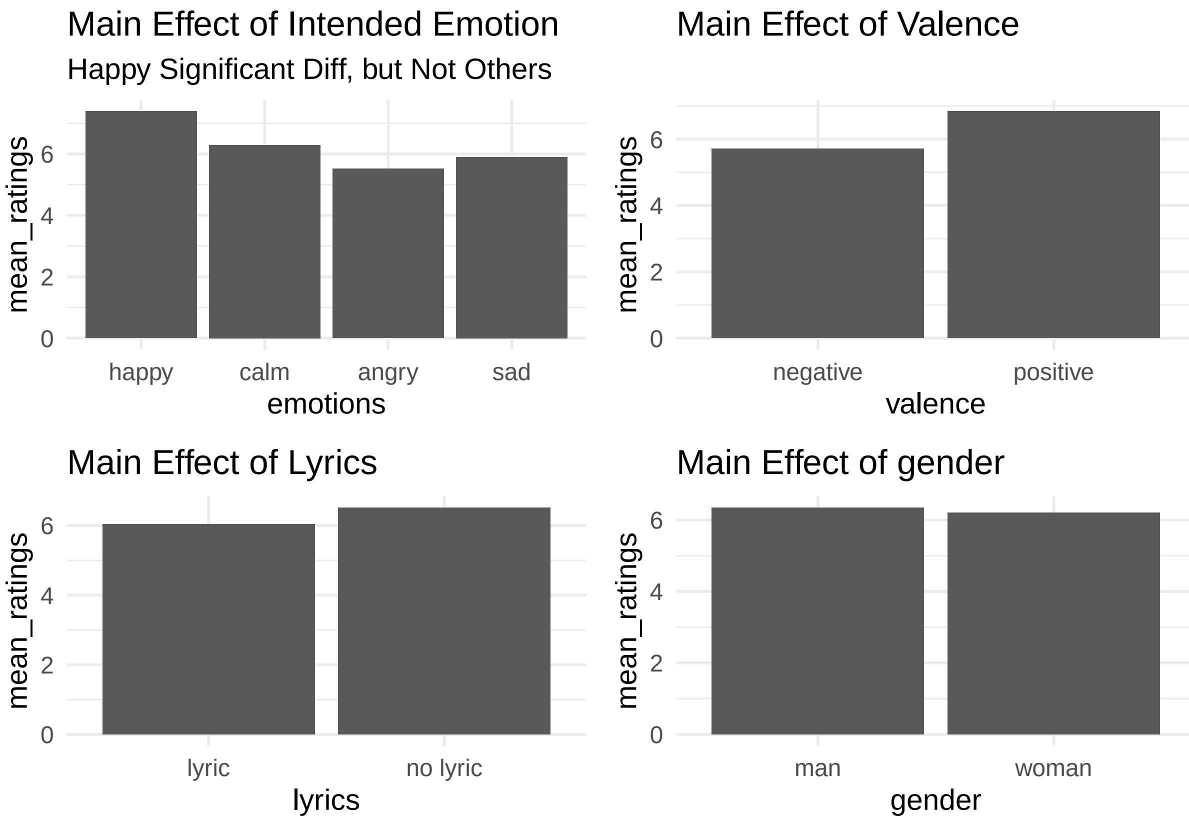
```
## Joining, by = "subject"

## `summarise()` regrouping output by 'Condition' (override with `.groups` argument)
```

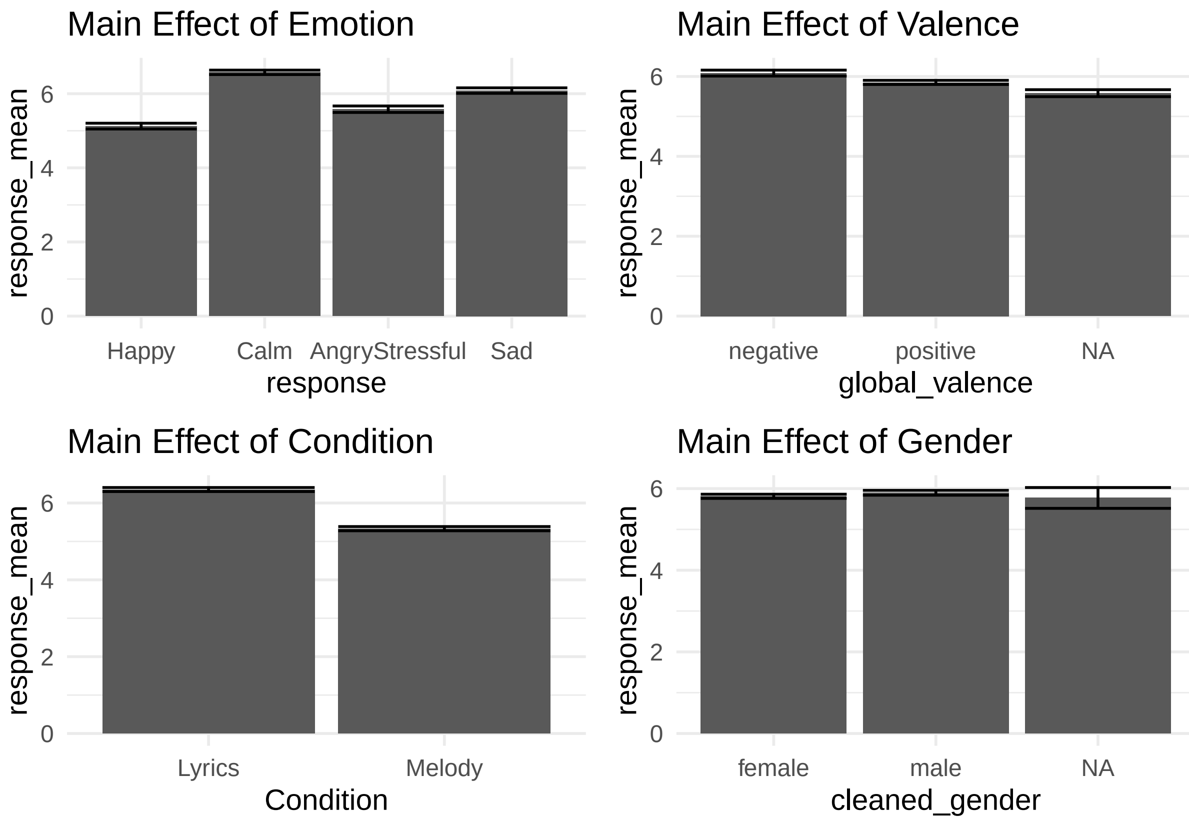



```r
# Big Plots of AP (Data Problem)

ap_panel
```


Main Effect of Intended Emotion
Happy Significant Diff, but Not Others
mean_ratings
0
2
4
6
happy
calm
angry
sad
emotions
Main Effect of Valence
mean_ratings
negative
positive
valence
Main Effect of Lyrics
mean_ratings
lyric
no lyric
lyrics
Main Effect of gender
mean_ratings
man
woman
gender

```
lsu_panel
```



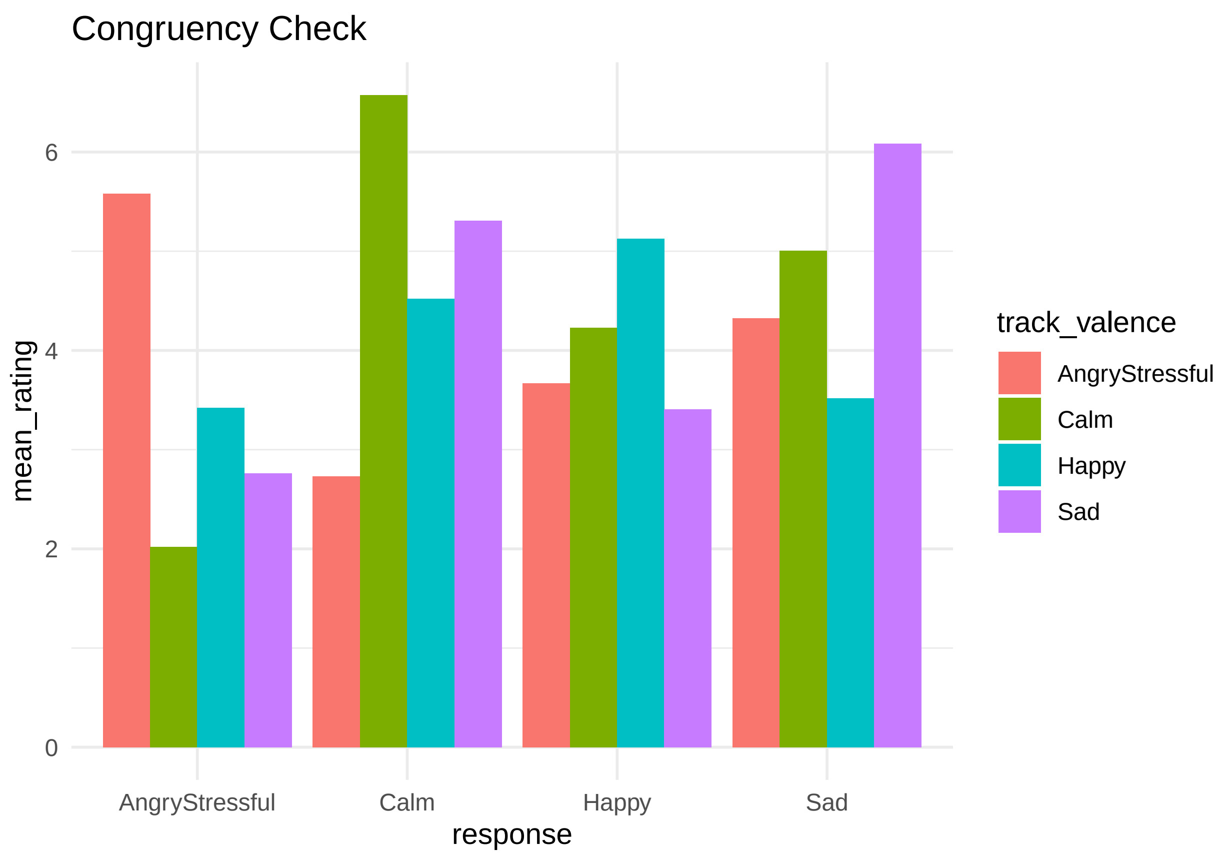
Before modeling, what do we see in comparision:

- We find happy is not highest, but calm, happy actually our lowest
- ANOVA to determine any sort of post-hoc significant
- We find reverse of valence, actually negative valence get higher ratings
- Appears people rate our lyrics as less, oppositive of others
- Also don't see any big evidence of main effect of gender
- Factorial ANOVA needed to Look at it all

Also wanted to see if congruancy always higher as they report but no numbers or graphs:

```
df_complete %>%
  select(subject, Happy:AngryStressful, Condition, track_valence, global_valence) %>%
  pivot_longer(cols = Happy:AngryStressful, names_to = "response",values_to = "rating") %>%
  group_by(track_valence, response) %>%
  summarise(
    mean_rating = mean(rating)
  ) %>%
  ggplot(aes(x = response, fill = track_valence, y = mean_rating)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Congruency Check")
```

```
## `summarise()` regrouping output by 'track_valence' (override with `.groups` argument)
```



Statistical Analysis

- First Re-Report the 4 x 2 x 2 ANOVA (Not with Gender)

```
options(scipen = 999)
df_replicate_anova <- df_congruent

factorial_anova_direct_replication <- ezANOVA(
  data = df_replicate_anova
  , dv = .(rating)
  , wid = .(subject)
  , within = .(Condition, response),
  type = 3, # Do you want to do Type III sum of Squares, doesnt change results...
  detailed = TRUE,
  return_aov = TRUE
)
```

```
## Warning: Converting "subject" to factor for ANOVA.
```

```
## Warning: Converting "Condition" to factor for ANOVA.
```

```
## Warning: Collapsing data to cell means. *IF* the requested effects are a subset
## of the full design, you must use the "within_full" argument, else results may be
## inaccurate.
```

```r
# Check Exlusion Criteria from Pre-Registration since this is on border
print(factorial_anova_direct_replication)
```

```
## $ANOVA
##              Effect DFn DFd          SSn      SSd           F
## 1       (Intercept)   1 128 35543.790940 847.8497 5366.051696
## 2         Condition   1 128   268.096899 124.1687  276.369133
## 3          response   3 384   261.893653 998.9657   33.557095
## 4 Condition:response   3 384     5.250484 262.3589    2.561613
##
## 1 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
## 2 0.0000000000000000000000000000000009041560394764533048728672327013248744772561004464030962488777516
## 3 0.0000000000000000002777326551014278800132935452982074364358195554580076248527009674660348537145182
## 4 0.0545824389638071144625897090918442700058221817016601562500000000000000000000000000000000000000000
##   p<.05         ges
## 1     * 0.940881089
## 2     * 0.107177029
## 3     * 0.104957440
## 4       0.002345439
##
## $`Mauchly's Test for Sphericity`
##              Effect         W           p p<.05
## 3          response 0.8553007 0.001358939     *
## 4 Condition:response 0.9462910 0.220972709
##
## $`Sphericity Corrections`
##              Effect       GGe                      p[GG] p[GG]<.05       HFe
## 3          response 0.8972105 0.000000000000000001486788         * 0.9183225
## 4 Condition:response 0.9622888 0.056926761167192016510620           0.9868555
##                      p[HF] p[HF]<.05
## 3 0.000000000000000006562326         *
## 4 0.055388119059531989996081
##
## $aov
##
## Call:
## aov(formula = formula(aov_formula), data = data)
##
## Grand Mean: 5.868702
##
## Stratum 1: subject
##
## Terms:
##                 Residuals
## Sum of Squares   847.8497
## Deg. of Freedom       128
##
## Residual standard error: 2.573679
##
## Stratum 2: subject:Condition
##
## Terms:
##                 Condition Residuals
```

```
## Sum of Squares   268.0969  124.1687
## Deg. of Freedom         1       128
##
## Residual standard error: 0.9849204
## 3 out of 4 effects not estimable
## Estimated effects are balanced
##
## Stratum 3: subject:response
##
## Terms:
##                  response Residuals
## Sum of Squares  261.8937  998.9657
## Deg. of Freedom        3       384
##
## Residual standard error: 1.612908
## 3 out of 6 effects not estimable
## Estimated effects may be unbalanced
##
## Stratum 4: subject:Condition:response
##
## Terms:
##                  Condition:response Residuals
## Sum of Squares             5.25048 262.35889
## Deg. of Freedom                  3       384
##
## Residual standard error: 0.826575
## Estimated effects may be unbalanced
```

- Main Effect of Condition (Lyrics) and Response (Emotion)
- No significant interaction with all data
- (it does emerge with missing gender people in there)

Data Caveat

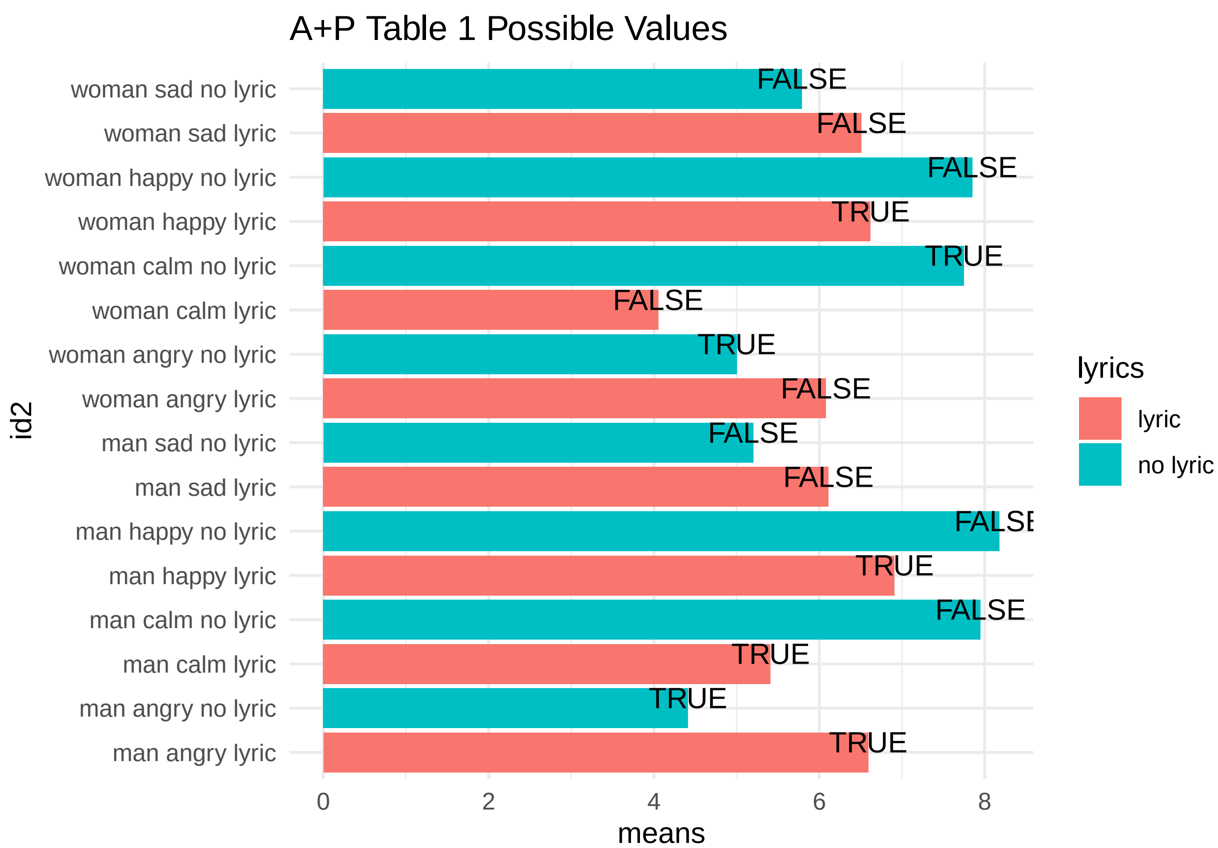
When making the top charts before running our data, I was kind of shocked at how low their sample size was. While that is not too much of a problem in general, I had just read this article and for fun wanted to see what would happen if I put their data into this calculator. Their preprint is here.

The long story short is that only seven out of their sixteen values were actually mathematically possible given their sample size. This could just be sloppieness on their part. But I looked into it a bit more and plotted which means were possible and which were not. What I found a bit peculiar, is the first main effect that they report of condition, with happy being higher than the other ones (but no difference between the two), the two values here that are the highest in the happy condition (women, happy, no lyric + man happy no lyric) are both impossible values. I didn't want to get into more speculation, but this probably should be addressed in the Discussion section.

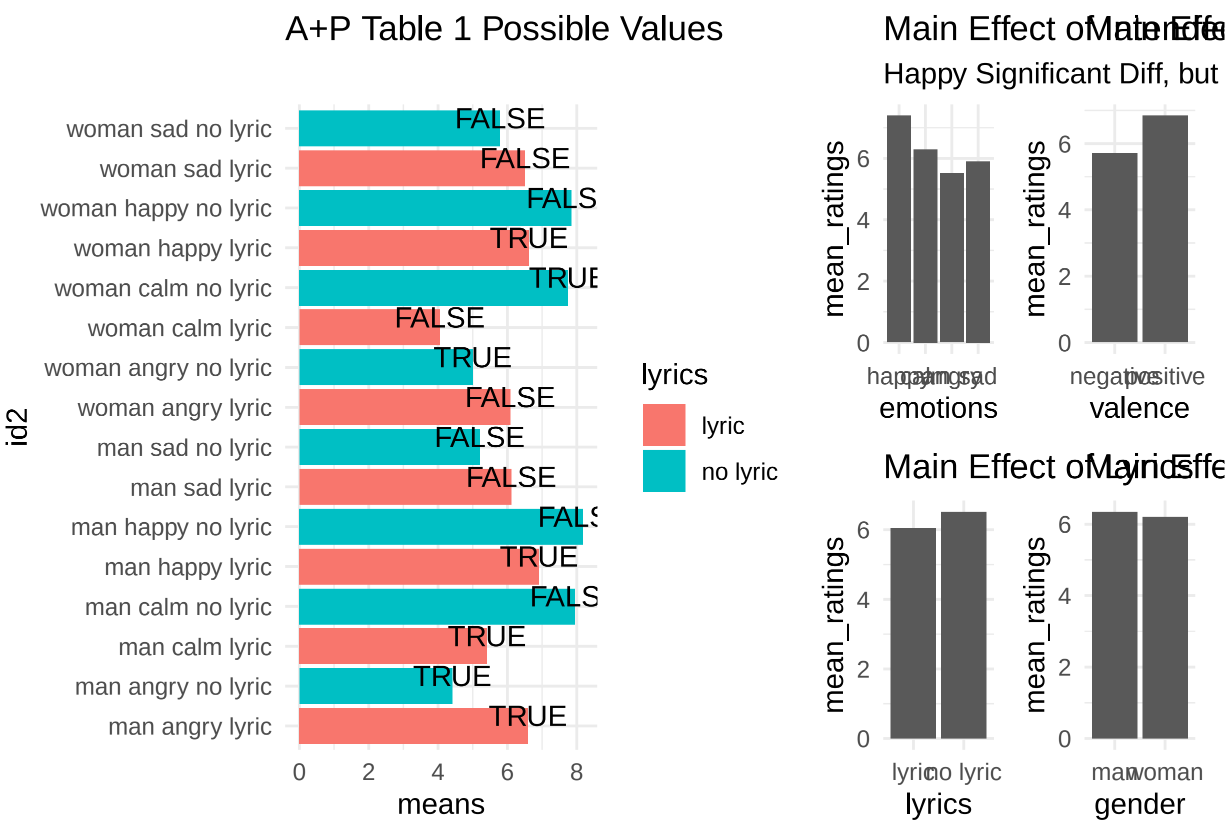
```r
table_1 %>%
  mutate(number_of_possible = sum(possible)) %>%
  mutate(total_values = length(possible)) %>%
  mutate(percent_possible = number_of_possible/total_values) %>%
  select(-valence, -number_of_possible, -total_values) %>%
  gt()
```

bottom	top	means	sample_size	possible	gender	lyrics	emotions	percent_possible
7.84	7.88	7.85	32	FALSE	woman	no lyric	happy	0.4375
8.16	8.19	8.18	32	FALSE	man	no lyric	happy	0.4375
6.59	6.62	6.62	32	TRUE	woman	lyric	happy	0.4375
6.91	6.94	6.91	32	TRUE	man	lyric	happy	0.4375
5.78	5.81	5.79	32	FALSE	woman	no lyric	sad	0.4375
5.19	5.22	5.20	32	FALSE	man	no lyric	sad	0.4375
6.50	6.53	6.51	32	FALSE	woman	lyric	sad	0.4375
6.09	6.12	6.11	32	FALSE	man	lyric	sad	0.4375
7.75	7.75	7.75	32	TRUE	woman	no lyric	calm	0.4375
7.94	7.97	7.95	32	FALSE	man	no lyric	calm	0.4375
4.03	4.06	4.05	32	FALSE	woman	lyric	calm	0.4375
5.41	5.44	5.41	32	TRUE	man	lyric	calm	0.4375
5.00	5.00	5.00	32	TRUE	woman	no lyric	angry	0.4375
4.41	4.44	4.41	32	TRUE	man	no lyric	angry	0.4375
6.06	6.09	6.08	32	FALSE	woman	lyric	angry	0.4375
6.56	6.59	6.59	32	TRUE	man	lyric	angry	0.4375

```
(possible_bar_chart <- table_1 %>%
  tibble() %>%
  mutate(id = row_number()) %>%
  mutate(id2 = paste(gender, emotions, lyrics)) %>%
  ggplot(aes(y = means,
             x = id2,
             fill  = lyrics)) +
  geom_bar(stat = "identity") +
  stat_summary(fun = sum, geom="text", aes(label=possible), vjust = 0) +
  theme_minimal() +
  coord_flip() +
  labs(title = "A+P Table 1 Possible Values") )
```




```
possible_bar_chart + ap_panel
```



Extending

I think we should write the paper and say that since they didn't find an effect of gender, it was not worth testing. They had some weird literature suggesting why it should be added, but in order to figure out if the results here are generlizable (obvs not since we didn't see any other patterns consistent with their original findings) we have other ideas that we might put forward.

Instead of thinking that there would be a main effect of gender, we might think that people with musical training will respond with more variability than those without.

To explore this we do the following

- Correlate GMSIs with mean rating in congruancy conditions
- Then add GMSI in linear model as interaction

  rating ~ emotions + lyrics:gmsi

Or somethingl like that.