Mixed Effects Write Up Draft

David John Baker

20/08/2020

This document provides the draft of the write up for the mixed effects analysis. It includes three analyses:

1. Use individual to see if inclusion of musical training affects results (Q35)
2. Create musical model to see if chord families are different
3. Investigate the extent that the five features contribute to ratings

Additionally, this analysis explores Table 7 which attempts to ascertain which scale degrees were most important in determining the attraction ratings.

The text below is written to be used in the final manuscript.

Data Import

Results

In order to analyze the results from our experiment, we ran three separate linear mixed effects models to investigate what variation in our data was attributable to our experimental conditions, what variation was attributable to within-subject variation, and what variance was left unexplained considering our hypothesized variables of interest. These included individual differences due to musical training, the chord family that the pre-dominant chord belonged as stated in TABLE X, and the chord's features as stated in TABLE X.

Prior to investigating the extent that each of our fixed effects contributed to the model, we first ran a null model that partitioned the within-subject effects due to participant in order to establish a baseline measure we could compare further models to. In this and all subsequent models, participant was modeled as a random effect with our dependent variable being the attraction rating. Doing this allowed for us to account for both the violation of the independence assumption since the same participant rated each stimuli twice, once in each block.

Subsequent to our null model where participant was modeled as a random effect, we then ran our first linear mixed effects model to estimate the extent that musical training affected our attraction ratings. Our second model included musical training as a paramter to also include and then finally in our third model we attempted to simultaneously consider both individual and musical features in order to examine the extent that the five features FROM THIS TABLE contributed to the attraction ratings. In each of these three models, in addition to allowing for random intercepts for participants– building on the null model– we allowed random slopes for each of the fixed effects. We chose not to run all variables simultaneously in order to prevent a singular model fit also known as over-fitting. Data analyses were run using the R programming language (CITE) using the `lme4` package (Bates et al). Both data and analyses are available at .

Null Model

We first present the results of our null model. Here we modelled our dependent variable of attraction using both participant and block and fully crossed random effects.

```
# Select Individual Model Data
model_data <- df %>%
  select(rating,
         chord_family,
         lerdhal_tension,
         parncut_roughness,
         semitone_voice_move,
         rootmotion,
         number_tendency_tones4_6,
         participant,
         block,
         i_have_had_formal_training_in_music_theory_for_years,
         f3_musical_training,experimental_group) %>%
  mutate(experimental_group = factor(experimental_group, c("prolific", "freshman","upperclass")))

# Run Null Model
null_model <- lmer(rating ~ (1|participant) + (1|block), data = model_data)
null_summary <- summary(null_model)
null_summary
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: rating ~ (1 | participant) + (1 | block)
##    Data: model_data
##
## REML criterion at convergence: 12841.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2894 -0.6047  0.0866  0.7197  3.2681
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  participant (Intercept) 0.4756497 0.68967
##  block       (Intercept) 0.0003334 0.01826
##  Residual                1.8710928 1.36788
## Number of obs: 3658, groups:  participant, 59; block, 2
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  4.64689    0.09349 44.22418   49.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(paste(round(.47/(.47+1.87)*100,3),"% of the data is explained with null model"))
```

```
## [1] "20.085 % of the data is explained with null model"
```

```
print("Variation due to block is negligable")
```

```
## [1] "Variation due to block is negligable"
```

Running the null model, we were able to account for 20% of the variation in response due solely to variation in response of participants. Of this 20%, a negligible amount was due to block from which we can conclude no effects of block. This variable was not used for subsequent analyses.

Model 1

After the null model, we then ran a model that that listed `chord_family` as a fixed effect. Since we theorized that there is something special about chord function/family, chord family was a fixed effect which allowed random intercepts.

```
# Chord with Random Intercept
chord_category_model <- lmer(rating ~ chord_family + (1+chord_family|participant), data = model_data)
summary(chord_category_model)
```

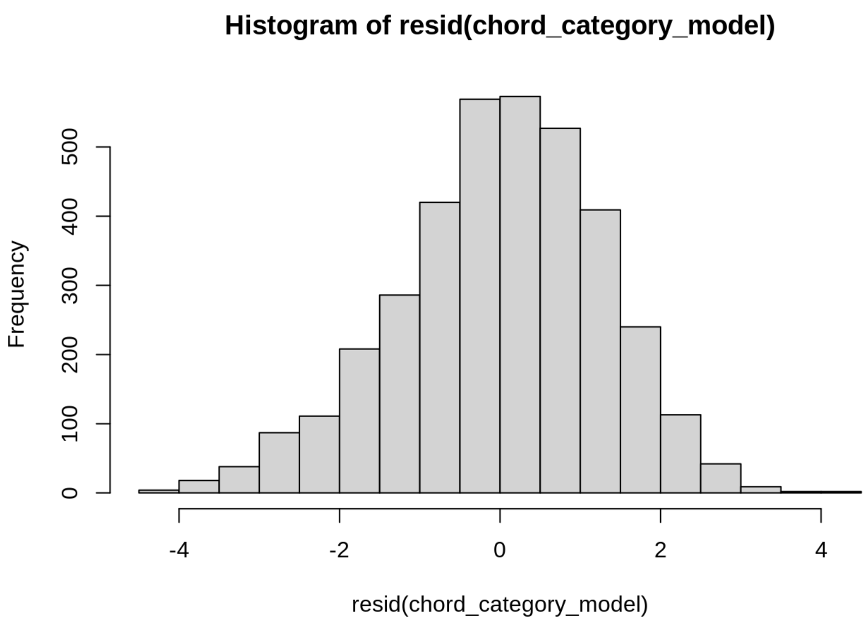
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: rating ~ chord_family + (1 + chord_family | participant)
##    Data: model_data
##
## REML criterion at convergence: 12613.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3616 -0.6063  0.0699  0.6857  3.1368
##
## Random effects:
##  Groups      Name                     Variance Std.Dev. Corr
##  participant (Intercept)              0.53121  0.7288
##              chord_familychromatic_pd 0.59516  0.7715   -0.36
##              chord_familycommon_pd    0.43252  0.6577   -0.37  0.86
##              chord_familynot_pd       0.03898  0.1974   -0.08  0.63  0.53
##  Residual                             1.69128  1.3005
## Number of obs: 3658, groups:  participant, 59
##
## Fixed effects:
##                          Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)               4.39518    0.10390 57.99998  42.302  < 2e-16 ***
## chord_familychromatic_pd  0.41810    0.11692 57.99874   3.576 0.000712 ***
## chord_familycommon_pd     0.52897    0.10447 57.99938   5.063 4.47e-06 ***
## chord_familynot_pd        0.03238    0.06708 58.00205   0.483 0.631127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) chrd_fmlych_ chrd_fmlycm_
## chrd_fmlych_ -0.427
## chrd_fmlycm_ -0.442  0.751
## chrd_fmlyn_  -0.285  0.437        0.421
```

```
# Significantly better with chord family in model
anova(null_model, chord_category_model)
```

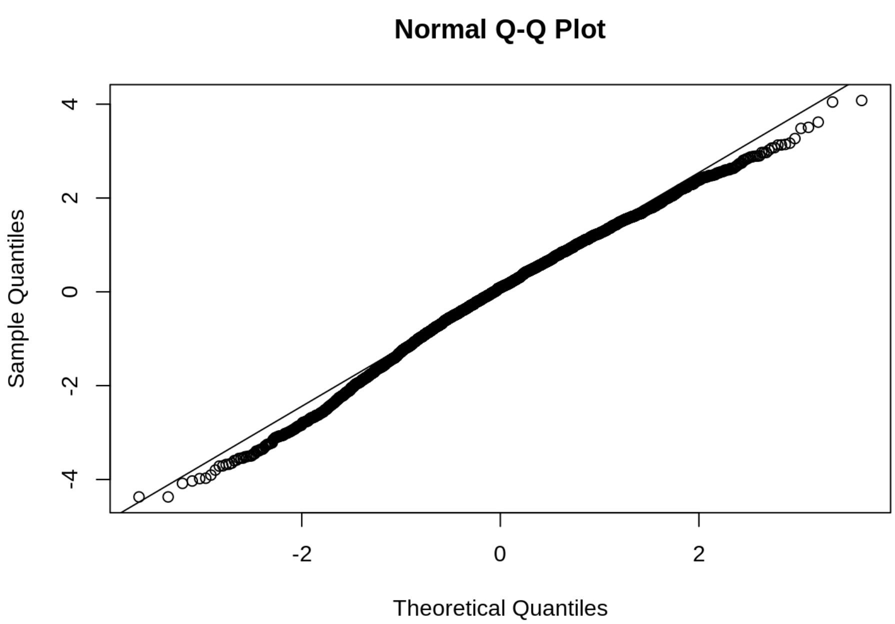
```
## refitting model(s) with ML (instead of REML)
```

```
## Data: model_data
## Models:
## null_model: rating ~ (1 | participant) + (1 | block)
## chord_category_model: rating ~ chord_family + (1 + chord_family | participant)
##                      npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## null_model              4 12846 12871 -6419.2    12838
## chord_category_model   15 12630 12723 -6300.2    12600 238.05 11  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Residuals were normally distributed
hist(resid(chord_category_model))
```




```
qqnorm(resid(chord_category_model))
qqline(resid(chord_category_model))
```




```
r.squaredGLMM(null_model)
```

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
##      R2m       R2c
## [1,]   0 0.2027983
```

```
r.squaredGLMM(chord_category_model)
```

```
##             R2m       R2c
## [1,] 0.02306407 0.2804441
```

```
library(sjPlot)
tab_model(chord_category_model,title = "Chord Category Model")
```

Chord Category Model

	rating		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.40	4.19 – 4.60	**<0.001**
chord_family [chromatic_pd]	0.42	0.19 – 0.65	**<0.001**
chord_family [common_pd]	0.53	0.32 – 0.73	**<0.001**
chord_family [not_pd]	0.03	-0.10 – 0.16	0.629
Random Effects			
σ^2	1.69		
$\tau_{00\ \text{participant}}$	0.53		
$\tau_{11\ \text{participant.chord\_familychromatic\_pd}}$	0.60		
$\tau_{11\ \text{participant.chord\_familycommon\_pd}}$	0.43		
$\tau_{11\ \text{participant.chord\_familynot\_pd}}$	0.04		
ρ_{01}	-0.36		
	-0.37		
	-0.08		
ICC	0.26		
$N_{\text{participant}}$	59		
Observations	3658		
Marginal R^2 / Conditional R^2	0.023 / 0.280		

Results from using `chord_family` as a fixed effect resulted in a signifcantly better model fit ($\chi^2(1$ $238.05, p < .001$). With chord family as a fixed effect, our marginal $R^2 = .02$, with our conditional $R^2 = .28$. Coeffecients from the model can be found in the Chord Category Model table or FIGURE X (below).

```
std <- function(x) sd(x)/sqrt(length(x))

# Chord Random Slopes
chord_category_model_random_slopes <- lmer(rating ~ chord_family + (1+ chord_family | participant), data = model_
data, REML = TRUE)
(ccmrs_summary <- summary(chord_category_model_random_slopes))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: rating ~ chord_family + (1 + chord_family | participant)
##    Data: model_data
##
## REML criterion at convergence: 12613.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3616 -0.6063  0.0699  0.6857  3.1368
##
## Random effects:
##  Groups      Name                     Variance Std.Dev. Corr
##  participant (Intercept)              0.53121  0.7288
##              chord_familychromatic_pd 0.59516  0.7715   -0.36
##              chord_familycommon_pd    0.43252  0.6577   -0.37  0.86
##              chord_familynot_pd       0.03898  0.1974   -0.08  0.63  0.53
##  Residual                             1.69128  1.3005
## Number of obs: 3658, groups:  participant, 59
##
## Fixed effects:
##                          Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)               4.39518    0.10390 57.99998  42.302  < 2e-16 ***
## chord_familychromatic_pd  0.41810    0.11692 57.99874   3.576 0.000712 ***
## chord_familycommon_pd     0.52897    0.10447 57.99938   5.063 4.47e-06 ***
## chord_familynot_pd        0.03238    0.06708 58.00205   0.483 0.631127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) chrd_fmlych_ chrd_fmlycm_
## chrd_fmlych_ -0.427
## chrd_fmlycm_ -0.442  0.751
## chrd_fmlyn_  -0.285  0.437        0.421
```

```
(.54+.0002+.38+.51)/(.54+.0002+.38+.51+1.69)
```

```
## [1] 0.4583681
```

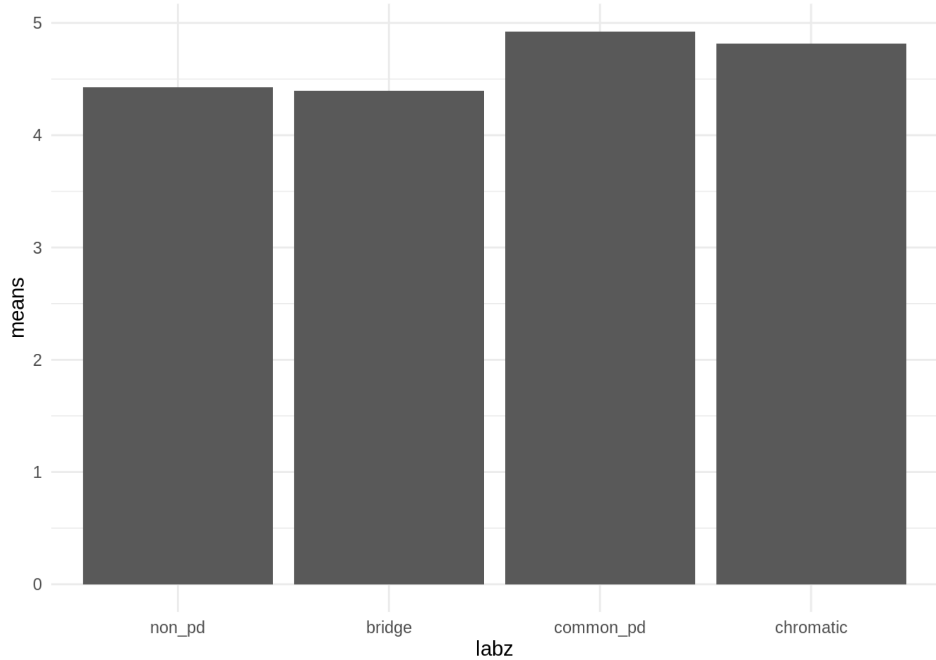
```
ccmrs_summary$coefficients[,1]
```

```
##              (Intercept) chord_familychromatic_pd    chord_familycommon_pd
##               4.39517585               0.41810169               0.52897034
##       chord_familynot_pd
##               0.03237984
```

```
means <- tibble(means = c(4.42755569,4.42755569+-0.0323784,4.42755569+.49659050,4.42755569+0.38572185),
                labz = c("non_pd","bridge","common_pd","chromatic"))

means %>%
  mutate(labz = factor(labz, levels = c("non_pd","bridge","common_pd","chromatic"))) %>%
  ggplot(aes(y = means, x = labz)) +
  geom_bar(stat = "identity") +
  theme_minimal()-> chord_family_model_plot_1

chord_family_model_plot_1
```



Model 2

Building on Model 1, we then retained our chord family variable as a fixed effect, categorical predictor and modeled participant as a random intercept and years of formal musical training (question 35 from the Goldsmiths Musical Sophistication Index) as a random slope. This allowed the model to account for the fact that ratings for chord attraction will vary as a result of how much formal, musical training an individual will recieve.

```
musical_individual_model <- lmer(rating ~ chord_family  +
                                   (1+ i_have_had_formal_training_in_music_theory_for_years |participant),
                                 data = model_data)

summary(musical_individual_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## rating ~ chord_family + (1 + i_have_had_formal_training_in_music_theory_for_years |
##     participant)
##    Data: model_data
##
## REML criterion at convergence: 12739.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1488 -0.6322  0.0572  0.7183  3.2243
##
## Random effects:
##  Groups      Name                                                 Variance
##  participant (Intercept)                                          1.30394
##              i_have_had_formal_training_in_music_theory_for_years 0.06763
##  Residual                                                         1.81769
##  Std.Dev. Corr
##  1.1419
##  0.2601   -0.91
##  1.3482
## Number of obs: 3658, groups:  participant, 59
##
## Fixed effects:
##                             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)               4.432e+00  9.112e-02 6.737e+01  48.640  < 2e-16 ***
## chord_familychromatic_pd 4.181e-01  6.206e-02 3.596e+03   6.737 1.87e-11 ***
## chord_familycommon_pd    5.290e-01  6.206e-02 3.596e+03   8.524  < 2e-16 ***
## chord_familynot_pd       3.238e-02  6.423e-02 3.596e+03   0.504    0.614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) chrd_fmlych_ chrd_fmlycm_
## chrd_fmlych_ -0.341
## chrd_fmlycm_ -0.341  0.500
## chrd_fmlyn_  -0.329  0.483        0.483
```

```
(1.30+0.06)/(1.30+0.06+1.81)
```

```
## [1] 0.4290221
```

```
r.squaredGLMM(chord_category_model)
```

```
##             R2m       R2c
## [1,] 0.02306407 0.2804441
```

```
r.squaredGLMM(musical_individual_model)
```

```
##             R2m       R2c
## [1,] 0.02302891 0.2278405
```

```
anova(chord_category_model_random_slopes, musical_individual_model)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: model_data
## Models:
## musical_individual_model: rating ~ chord_family + (1 + i_have_had_formal_training_in_music_theory_for_years |
## musical_individual_model:     participant)
## chord_category_model_random_slopes: rating ~ chord_family + (1 + chord_family | participant)
##                                    npar   AIC   BIC  logLik deviance Chisq Df
## musical_individual_model              8 12740 12790 -6362.3    12724
## chord_category_model_random_slopes   15 12630 12723 -6300.2    12600 124.2  7
##                                    Pr(>Chisq)
## musical_individual_model
## chord_category_model_random_slopes  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tab_model(musical_individual_model, title ="Musical Training Added")
```

Musical Training Added

	rating		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.43	4.25 – 4.61	**<0.001**
chord_family [chromatic_pd]	0.42	0.30 – 0.54	**<0.001**
chord_family [common_pd]	0.53	0.41 – 0.65	**<0.001**
chord_family [not_pd]	0.03	-0.09 – 0.16	0.614
Random Effects			
σ^2	1.82		
τ_{00} participant	1.30		
τ_{11} participant.i_have_had_formal_training_in_music_theory_for_years	0.07		
ρ_{01} participant	-0.91		
ICC	0.42		
N participant	59		
Observations	3658		
Marginal R^2 / Conditional R^2	0.017 / 0.428		

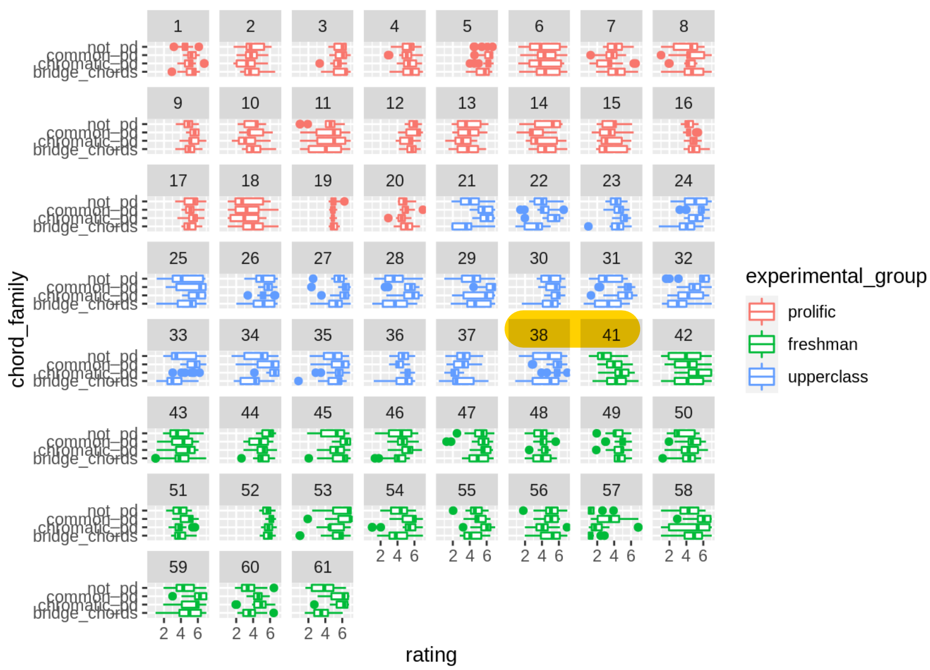
```
tab_model(chord_category_model, musical_individual_model)
```

	rating			rating		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	4.40	4.19 – 4.60	**<0.001**	4.43	4.25 – 4.61	**<0.001**
chord_family [chromatic_pd]	0.42	0.19 – 0.65	**<0.001**	0.42	0.30 – 0.54	**<0.001**
chord_family [common_pd]	0.53	0.32 – 0.73	**<0.001**	0.53	0.41 – 0.65	**<0.001**
chord_family [not_pd]	0.03	-0.10 – 0.16	0.629	0.03	-0.09 – 0.16	0.614
Random Effects						
σ^2	1.69			1.82		
τ_{00}	0.53 participant			1.30 participant		
τ_{11}	0.60 participant.chord_familychromatic_pd 0.43 participant.chord_familycommon_pd 0.04 participant.chord_familynot_pd			0.07 participant.i_have_had_formal_training_in_music_theory_for_years		
ρ_{01}	-0.36 -0.37 -0.08			-0.91 participant		
ICC	0.26			0.42		
N	59 participant			59 participant		
Observations	3658			3658		

Marginal R^2 / Conditional R^2	0.023 / 0.280	0.017 / 0.428

As evident from “MUSICAL TRAING ADDED TABLE”, while the coeffecients associated with chord family did not appear to change with the addition of this random slope parameter, the model did fit the data significantly better (χ^2(1 124.2, $p < .001$) and our marginal R^2 rose from .28 to .42, suggesting that while this information did help explain more of the data, there is variation due training that can be captured beyond using participant soley as a random intercept.

```
# Re-Do This with ordered Data on musical training ???
model_data %>%
  ggplot(aes(y = rating, x = chord_family, color = experimental_group)) +
  geom_boxplot() +
  facet_wrap(~participant) +
  coord_flip()
```




```
#ggsave("img/all_participants_data.png", width = 40, height = 20, units = "cm")
```

- PLOT HERE NOTE– DJB check after Round 1

Model 3

Finally, we then attempted to investigate the extent that the features from TABLE X contributed to attraction ratings. We now include all five features from TABLE X as fixed effects predictors, preserving the rest of the model structure from the prior models.

```
all_musical_indv_theory_model <- lmer(rating ~
                               chord_family +
                            lerdhal_tension +
                            parncut_roughness +
                            semitone_voice_move +
                            rootmotion +
                            number_tendency_tones4_6 +
                            (1+ i_have_had_formal_training_in_music_theory_for_years|participant),
                          data = model_data)

summary(all_musical_indv_theory_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: rating ~ chord_family + lerdhal_tension + parncut_roughness +
##     semitone_voice_move + rootmotion + number_tendency_tones4_6 +
##     (1 + i_have_had_formal_training_in_music_theory_for_years |
##         participant)
##    Data: model_data
##
## REML criterion at convergence: 11475.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0058 -0.6209  0.0349  0.7118  3.2313
##
## Random effects:
##  Groups      Name                                                Variance
##  participant (Intercept)                                         1.28017
##              i_have_had_formal_training_in_music_theory_for_years 0.07159
##  Residual                                                        1.78732
##  Std.Dev. Corr
##  1.1314
##  0.2676   -0.92
##  1.3369
## Number of obs: 3304, groups:  participant, 59
##
## Fixed effects:
##                           Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)                4.37880    0.23986 1821.85865  18.256  < 2e-16 ***
## chord_familychromatic_pd  -0.19453    0.11742 3237.00149  -1.657  0.09767 .
## chord_familycommon_pd      0.06195    0.10896 3237.00149   0.568  0.56974
## chord_familynot_pd         0.15056    0.10839 3237.00148   1.389  0.16491
## lerdhal_tension            0.04053    0.02725 3237.00148   1.487  0.13699
## parncut_roughness         -2.80605    0.34796 3237.00149  -8.064 1.03e-15 ***
## semitone_voice_move        0.15635    0.03097 3237.00149   5.049 4.68e-07 ***
## rootmotion                 0.41511    0.08238 3237.00149   5.039 4.94e-07 ***
## number_tendency_tones4_6   0.15216    0.05386 3237.00149   2.825  0.00476 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) chrd_fmlych_ chrd_fmlycm_ chrd_fmlyn_ lrdhl_ prnct_ smtn__
## chrd_fmlych_  0.240
## chrd_fmlycm_  0.133  0.797
## chrd_fmlyn_  -0.774 -0.023        0.067
## lerdhl_tnsn  -0.885 -0.234       -0.127        0.789
## prnct_rghns  -0.466  0.082        0.039        0.277       0.315
## semtn_vc_mv   0.183 -0.331       -0.168       -0.255      -0.287 -0.417
## rootmotion   -0.388 -0.384       -0.357        0.297       0.389  0.061  0.239
## nmbr_tn_4_6  -0.296 -0.741       -0.750        0.158       0.203  0.039  0.034
##             rotmtn
## chrd_fmlych_
## chrd_fmlycm_
## chrd_fmlyn_
## lerdhl_tnsn
## prnct_rghns
## semtn_vc_mv
## rootmotion
## nmbr_tn_4_6   0.115
```

```
r.squaredGLMM(all_musical_indv_theory_model)
```

```
##            R2m       R2c
## [1,] 0.0425073 0.2453274
```

Results from all three models can be seen in the table below.

```
tab_model(chord_category_model,musical_individual_model, all_musical_indv_theory_model)
```

	rating			rating			rating		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	
(Intercept)	4.40	4.19 – 4.60	**<0.001**	4.43	4.25 – 4.61	**<0.001**	4.38	3.91 – 4.85	**<0**

chord_family [chromatic_pd]	0.42	0.19 – 0.65	**<0.001**	0.42	0.30 – 0.54	**<0.001**	-0.19	-0.42 – 0.04	0
chord_family [common_pd]	0.53	0.32 – 0.73	**<0.001**	0.53	0.41 – 0.65	**<0.001**	0.06	-0.15 – 0.28	0
chord_family [not_pd]	0.03	-0.10 – 0.16	0.629	0.03	-0.09 – 0.16	0.614	0.15	-0.06 – 0.36	0
lerdhal_tension							0.04	-0.01 – 0.09	0
parncut_roughness							-2.81	-3.49 – -2.12	**<0**
semitone_voice_move							0.16	0.10 – 0.22	**<0**
rootmotion							0.42	0.25 – 0.58	**<0**
number_tendency_tones4_6							0.15	0.05 – 0.26	**0**
Random Effects									
σ^2	1.69			1.82			1.79		
τ_{00}	0.53 participant			1.30 participant			1.28 participant		
τ_{11}	0.60 participant.chord_familychromatic_pd 0.43 participant.chord_familycommon_pd 0.04 participant.chord_familynot_pd			0.07 participant.i_have_had_formal_training_in_music_theory_for_years			0.07 participant.i_have_had_formal_training_in_music_theory		
ρ_{01}	-0.36 -0.37 -0.08			-0.91 participant			-0.92 participant		
ICC	0.26			0.42			0.42		
N	59 participant			59 participant			59 participant		
Observations	3658			3658			3304		
Marginal R^2 / Conditional R^2	0.023 / 0.280			0.017 / 0.428			0.032 / 0.436		

From this table, it appears that while model the model was able to continue the trend in an albeit small direction of increasing both conditional and marginal R2 values, these increases might not lead to practical meaning or application. What is more of interest is the unstablizing of the chord family coeffecients when the features of the model are included. We follow up on this finding in CHORD SCALE DEGREE ANALYSIS (after discussing this analysis)

Discussion Points