E4 categorization

R. Cassano-Coleman

2025-10-02

This notebook analyzes categorization using Bayesian binomial generalized linear mixed effects models (GLMMs).

Set up

```
set.seed(15000)
```

```
data <- read_csv('../data/E1-E2-E4/categorization.csv', show_col_types = FALSE)
```

Convert variables to factors.

```
data %<>%
  mutate(scramble = ifelse(scramble == 'intact', 'Intact', scramble)) %>%
  mutate(exp_subject_id = as.factor(exp_subject_id),
         response = ifelse(response == 'Correct', TRUE, FALSE),
         scramble = factor(scramble, levels = c('Intact', '8B', '2B', '1B')),
         Musician = factor(Musician, levels = c('Yes', 'No'))) %>%
  filter(!is.na(response))
```

Set the contrast for condition.

```
contrasts(data$scramble) <- contr.treatment(4)
print(contrasts(data$scramble))
```

```
##        2 3 4
## Intact 0 0 0
## 8B     1 0 0
## 2B     0 1 0
## 1B     0 0 1
```

Set the musician/non-musician contrast.

```
contrasts(data$Musician) <- c(-1,1)
print(contrasts(data$Musician))
```

```
##     [,1]
## Yes   -1
## No     1
```

Main analysis

Priors

Priors are expressed in log(odds) space.

Intercept: Given that chance is 25%, we assume that participants will perform somewhere between chance and ceiling. We expect the center of the distribution of accuracy to be somewhere around 60% or 65%. If we use a center of 65% and an SD of 1.5, 95% of the values fall between 8.46% and 97.4%.

```
prior_intercept <- set_prior('normal(log(0.65 / (1 - 0.65)), 1)', class = 'Intercept')
```

Group: We might expect musicians to do slightly better than non-musicians, on average.

In this range, a difference in 0.25 log odds gives us about a 5% decrease in accuracy.

```
prior_mus <- set_prior('normal(-0.25, 1)', coef = 'Musician1')
```

Scramble: We expect performance to improve as scramble level decreases. If we code 8B as reference level, then we expect 8B > 2B and 8B > 1B.

Since we're keeping the musician slope at SD = 1, we'll keep these (and the interactions) at SD = 1. This seems to be a pretty weak prior.

```
prior_scramble8B <- set_prior('normal(-0.1, 1)', coef = 'scramble2')
prior_scramble2B <- set_prior('normal(-0.2, 1)', coef = 'scramble3')
prior_scramble1B <- set_prior('normal(-0.3, 1)', coef = 'scramble4')
```

Interaction: We expect no interaction between group and scramble.

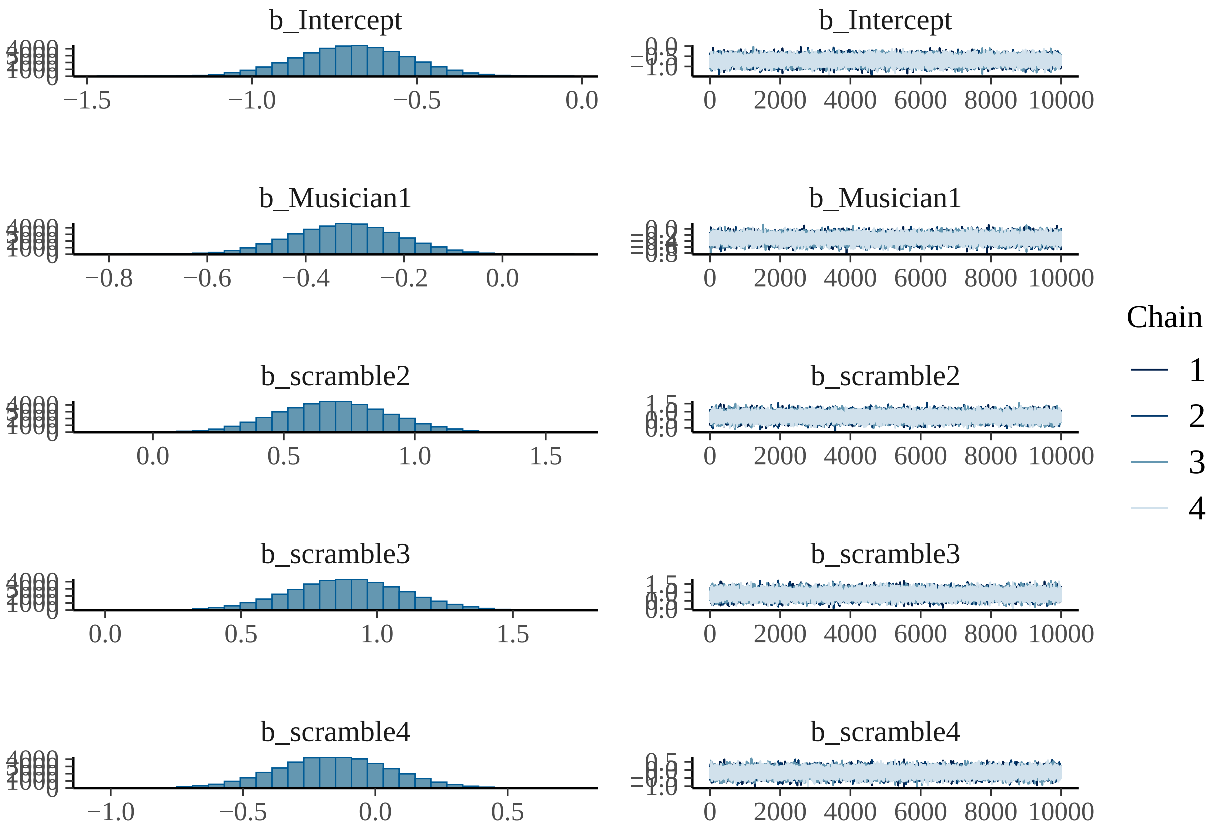
```
prior_int8B <- set_prior('normal(0, 1)', coef = 'Musician1:scramble2')
prior_int2B <- set_prior('normal(0, 1)', coef = 'Musician1:scramble3')
prior_int1B <- set_prior('normal(0, 1)', coef = 'Musician1:scramble4')
```

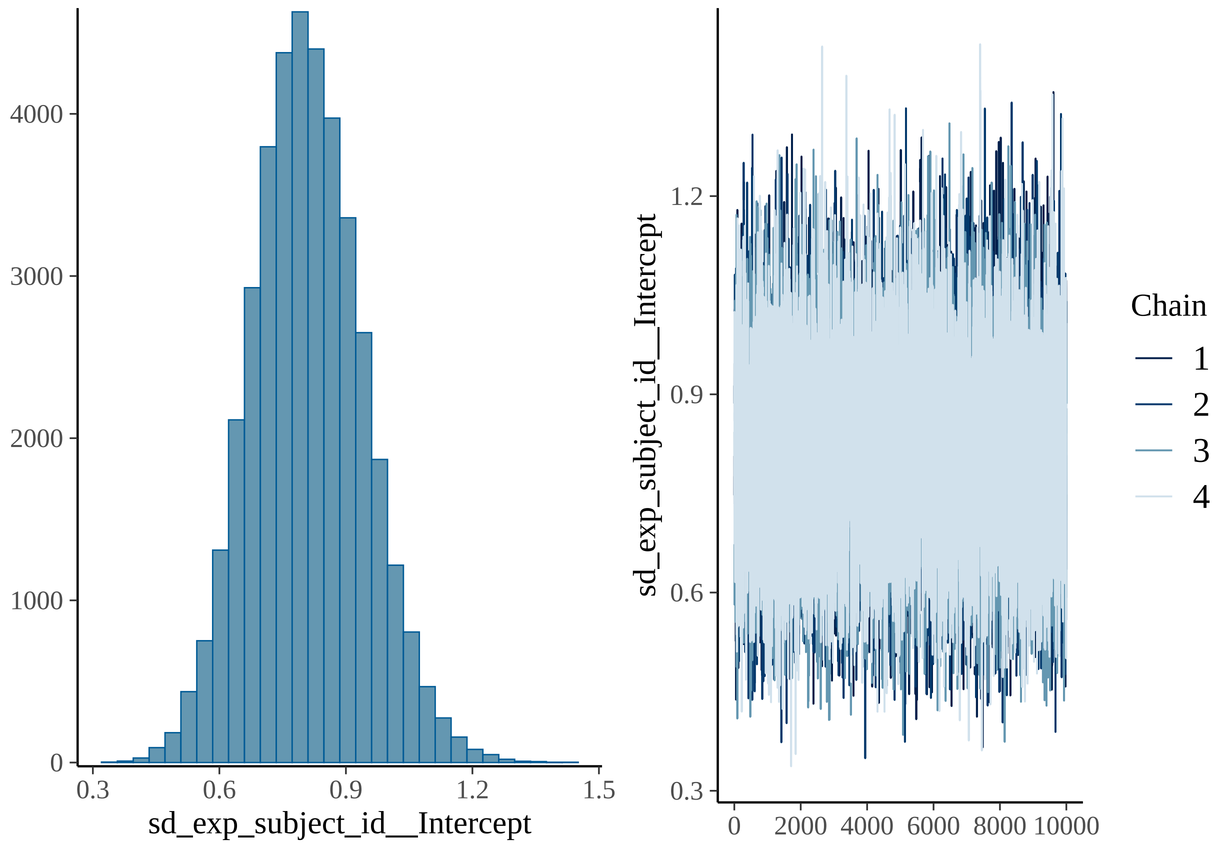
Random slope for subjects: *Leave this as default for now, may update.*

Main model with group and condition

```r
mus_scram <- brm(response ~ Musician + scramble + (1 | exp_subject_id), data = data,
                 family = bernoulli(),
                 prior = c(prior_intercept, prior_mus,
                           prior_scramble8B, prior_scramble2B, prior_scramble1B),
                 save_pars = save_pars(all = TRUE), iter = 20000, refresh = 0,
                 file = 'models/E4_mus_scram')
```

```r
plot(mus_scram)
```






```r
print(summary(mus_scram), digits = 4)
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: response ~ Musician + scramble + (1 | exp_subject_id)
##    Data: data (Number of observations: 818)
##   Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
##          total post-warmup draws = 40000
##
## Multilevel Hyperparameters:
## ~exp_subject_id (Number of levels: 106)
##               Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sd(Intercept)   0.8094    0.1315   0.5621   1.0784 1.0000    15611    21521
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## Intercept  -0.6965    0.1712  -1.0364  -0.3652 1.0001    51118    29845
## Musician1  -0.3196    0.1118  -0.5391  -0.1033 1.0000    43053    31346
## scramble2   0.6903    0.2122   0.2813   1.1086 1.0002    54540    33660
## scramble3   0.8910    0.2131   0.4756   1.3106 1.0000    56685    34399
## scramble4  -0.1566    0.2169  -0.5805   0.2691 1.0001    60528    32682
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
emm_mus_scram_s <- emmeans(mus_scram, specs = "scramble")
summary(emm_mus_scram_s)
```

```
##  scramble   emmean lower.HPD upper.HPD
##  Intact   -0.69491    -1.043    -0.373
##  8B       -0.00553    -0.346     0.321
##  2B        0.19436    -0.136     0.532
##  1B       -0.85085    -1.199    -0.495
##
## Results are averaged over the levels of: Musician
## Point estimate displayed: median
## Results are given on the logit (not the response) scale.
## HPD interval probability: 0.95
```

```r
contrast(emm_mus_scram_s, method = "pairwise")
```

```
##  contrast    estimate lower.HPD upper.HPD
##  Intact - 8B   -0.689    -1.104    -0.278
##  Intact - 2B   -0.890    -1.308    -0.474
##  Intact - 1B    0.157    -0.273     0.576
##  8B - 2B       -0.201    -0.613     0.220
##  8B - 1B        0.845     0.425     1.290
##  2B - 1B        1.048     0.607     1.474
##
## Results are averaged over the levels of: Musician
## Point estimate displayed: median
## Results are given on the log odds ratio (not the response) scale.
## HPD interval probability: 0.95
```

```r
emm_mus_scram_ms <- emmeans(mus_scram, specs = c("Musician", "scramble"))
summary(emm_mus_scram_ms)
```

```
##  Musician scramble emmean lower.HPD upper.HPD
##  Yes      Intact   -0.376   -0.7643    0.0198
##  No       Intact   -1.013   -1.4179   -0.5957
##  Yes      8B        0.313   -0.0825    0.7125
##  No       8B       -0.326   -0.7282    0.0752
##  Yes      2B        0.513    0.1238    0.9195
##  No       2B       -0.124   -0.5333    0.2684
##  Yes      1B       -0.532   -0.9355   -0.1267
##  No       1B       -1.170   -1.5976   -0.7475
##
## Point estimate displayed: median
## Results are given on the logit (not the response) scale.
## HPD interval probability: 0.95
```

```r
contrast(emm_mus_scram_ms, method = "pairwise")
```

```
##  contrast               estimate lower.HPD upper.HPD
##  Yes Intact - No Intact   0.6369     0.206     1.077
##  Yes Intact - Yes 8B     -0.6892    -1.104    -0.278
##  Yes Intact - No 8B      -0.0491    -0.650     0.536
##  Yes Intact - Yes 2B     -0.8905    -1.308    -0.474
##  Yes Intact - No 2B      -0.2514    -0.857     0.330
##  Yes Intact - Yes 1B      0.1571    -0.273     0.576
##  Yes Intact - No 1B       0.7962     0.180     1.403
```

```
##  No Intact - Yes 8B      -1.3258    -1.937    -0.716
##  No Intact - No 8B       -0.6892    -1.104    -0.278
##  No Intact - Yes 2B      -1.5252    -2.152    -0.921
##  No Intact - No 2B       -0.8905    -1.308    -0.474
##  No Intact - Yes 1B      -0.4829    -1.113     0.108
##  No Intact - No 1B        0.1571    -0.273     0.576
##  Yes 8B - No 8B           0.6369     0.206     1.077
##  Yes 8B - Yes 2B         -0.2009    -0.613     0.220
##  Yes 8B - No 2B           0.4381    -0.175     1.039
##  Yes 8B - Yes 1B          0.8452     0.425     1.290
##  Yes 8B - No 1B           1.4832     0.860     2.110
##  No 8B - Yes 2B          -0.8374    -1.455    -0.249
##  No 8B - No 2B           -0.2009    -0.613     0.220
##  No 8B - Yes 1B           0.2070    -0.405     0.805
##  No 8B - No 1B            0.8452     0.425     1.290
##  Yes 2B - No 2B           0.6369     0.206     1.077
##  Yes 2B - Yes 1B          1.0478     0.607     1.474
##  Yes 2B - No 1B           1.6817     1.050     2.302
##  No 2B - Yes 1B           0.4086    -0.199     1.012
##  No 2B - No 1B            1.0478     0.607     1.474
##  Yes 1B - No 1B           0.6369     0.206     1.077
##
## Point estimate displayed: median
## Results are given on the log odds ratio (not the response) scale.
## HPD interval probability: 0.95
```

Main effects

```r
main_BF <- describe_posterior(mus_scram,
                              estimate = "median", dispersion = TRUE,
                              ci = .95, ci_method = "HDI",
                              test = c("bayes_factor"))
print(main_BF, digits = 4)
```

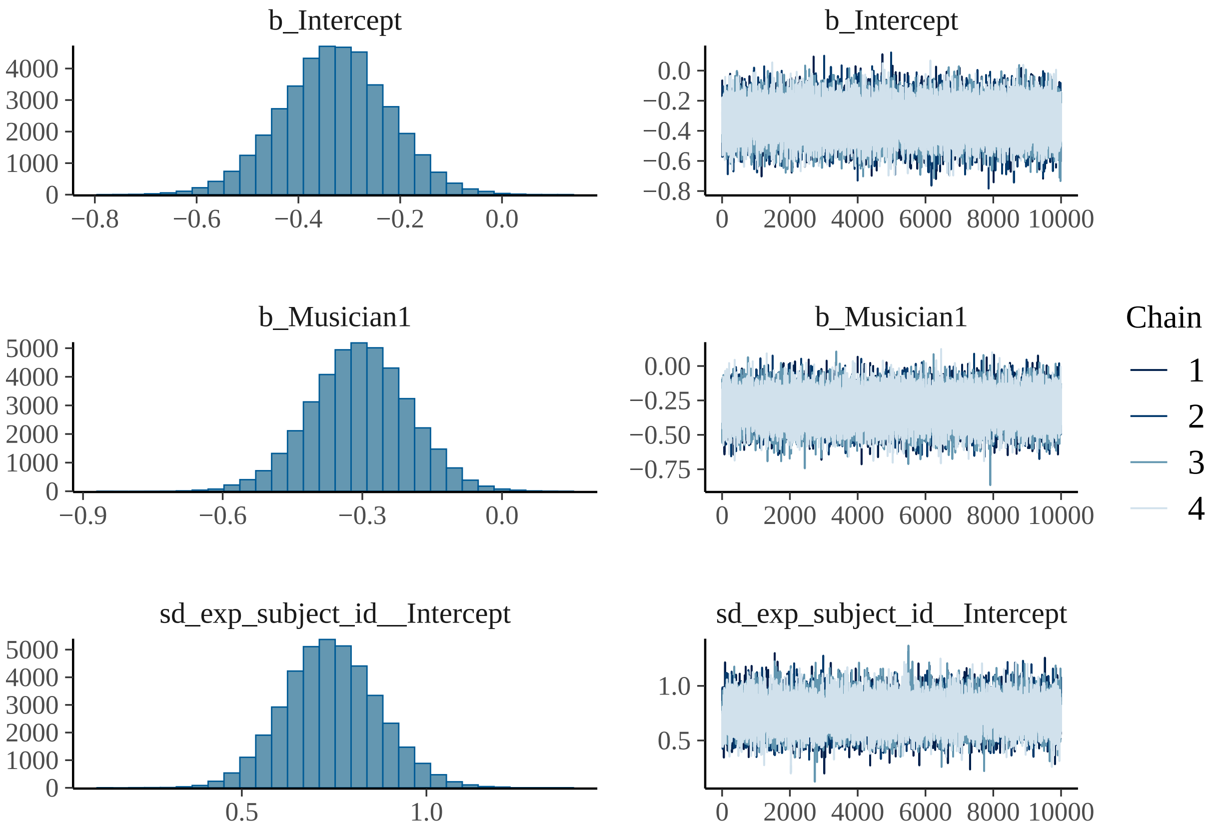
```
## Summary of Posterior Distribution
##
## Parameter   |  Median |    MAD |         95% CI |       BF |  Rhat |        ESS
## -------------------------------------------------------------------------------
## (Intercept) | -0.6949 | 0.1712 | [-1.04, -0.37] |   795.39 | 1.000 | 50997.0000
## Musician1   | -0.3185 | 0.1117 | [-0.54, -0.10] |     6.91 | 1.000 | 42978.0000
## scramble2   |  0.6892 | 0.2139 | [ 0.28,  1.10] |    50.54 | 1.000 | 54437.0000
## scramble3   |  0.8905 | 0.2122 | [ 0.47,  1.31] | 1.15e+03 | 1.000 | 56586.0000
## scramble4   | -0.1571 | 0.2161 | [-0.58,  0.27] |    0.266 | 1.000 | 60477.0000
```

Moderate evidence for a main effect of group.

To get the main effect of scramble level, fit the “null” model with group only to compare.

```
mus_only <- brm(response ~ Musician + (1 | exp_subject_id), data = data,
                 family = bernoulli(),
                 prior = c(prior_intercept, prior_mus),
                 save_pars = save_pars(all = TRUE), iter = 20000, refresh = 0,
                 file = 'models/E4_mus_only')
```

```
plot(mus_only)
```




```
print(summary(mus_only), digits = 4)
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: response ~ Musician + (1 | exp_subject_id)
##    Data: data (Number of observations: 818)
##   Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
##          total post-warmup draws = 40000
##
## Multilevel Hyperparameters:
## ~exp_subject_id (Number of levels: 106)
##               Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sd(Intercept)   0.7434    0.1283   0.4996   1.0035 1.0008    14297    22399
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## Intercept  -0.3281    0.1055  -0.5376  -0.1240 1.0000    39158    30512
## Musician1  -0.3054    0.1053  -0.5148  -0.1002 1.0000    39893    31391
##
```

```
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
BF_scramble <- bayes_factor(mus_scram, mus_only)
```

```
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
```

```r
print(BF_scramble)
```

```
## Estimated Bayes factor in favor of mus_scram over mus_only: 162016.64193
```

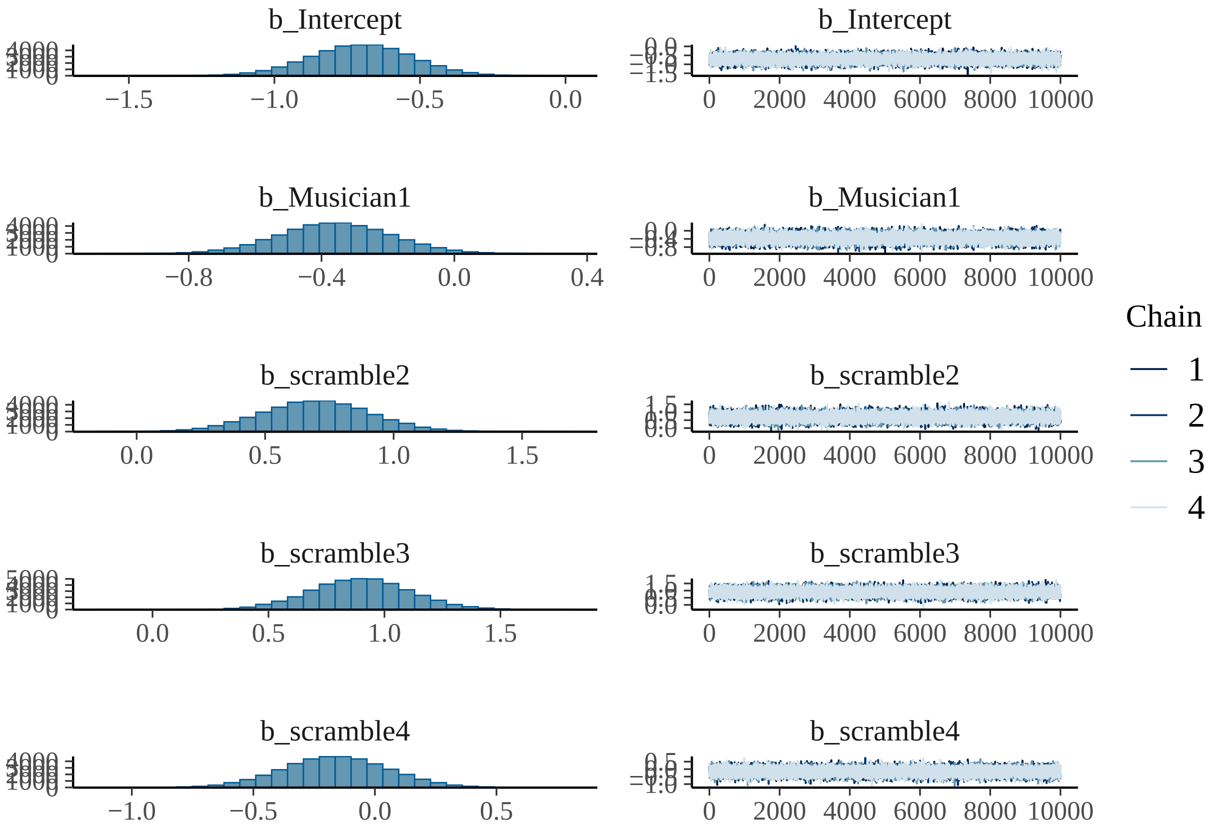
Very strong evidence for a main effect of scramble condition.

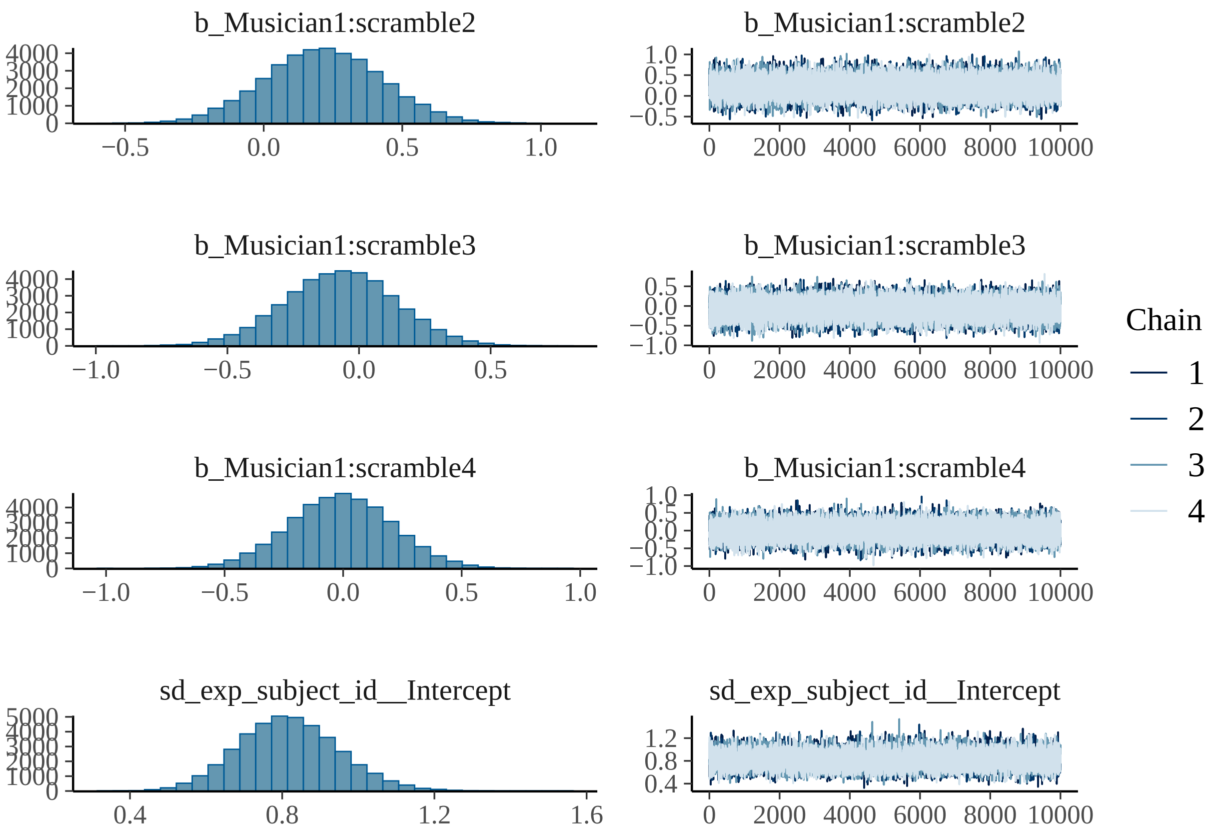
Interaction between group and condition?

Add an interaction between group and condition, and compare the model with the interaction to the one without.

```
mus_scram_int <- brm(response ~ Musician*scramble + (1 | exp_subject_id), data = data,
                  family = bernoulli(),
                  prior = c(prior_intercept, prior_mus,
                            prior_scramble8B, prior_scramble2B, prior_scramble1B,
                            prior_int8B, prior_int2B, prior_int1B),
                  save_pars = save_pars(all = TRUE), iter = 20000, refresh = 0,
                  file = 'models/E4_mus_scram_int')
```

```
plot(mus_scram_int)
```






```r
print(summary(mus_scram_int), digits = 4)
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: response ~ Musician * scramble + (1 | exp_subject_id)
##    Data: data (Number of observations: 818)
##   Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
##          total post-warmup draws = 40000
##
## Multilevel Hyperparameters:
## ~exp_subject_id (Number of levels: 106)
##               Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sd(Intercept)   0.8186    0.1322   0.5710   1.0895 1.0001    16177    22702
##
## Regression Coefficients:
##                     Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS
## Intercept            -0.7028    0.1737  -1.0481  -0.3656 1.0001    45376
## Musician1            -0.3585    0.1699  -0.6943  -0.0256 1.0002    37265
## scramble2             0.6984    0.2124   0.2828   1.1148 1.0002    54858
## scramble3             0.8963    0.2144   0.4783   1.3192 1.0001    53643
## scramble4            -0.1617    0.2168  -0.5851   0.2612 1.0000    56725
## Musician1:scramble2   0.2158    0.2114  -0.1949   0.6323 1.0000    44910
## Musician1:scramble3  -0.0698    0.2119  -0.4918   0.3445 1.0000    45835
## Musician1:scramble4  -0.0092    0.2181  -0.4375   0.4197 0.9999    45722
##                     Tail_ESS
## Intercept              29721
## Musician1              31033
## scramble2              33867
```

```
## scramble3              32999
## scramble4              34500
## Musician1:scramble2    32434
## Musician1:scramble3    33276
## Musician1:scramble4    33262
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
BF_int <- bayes_factor(mus_scram_int, mus_scram)
```

```
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
```

```r
print(BF_int)
```

```
## Estimated Bayes factor in favor of mus_scram_int over mus_scram: 0.01964
```

Strong evidence against an interaction between group and condition.

Figure 5

Create a helper function for the conversion from log odds to probability.

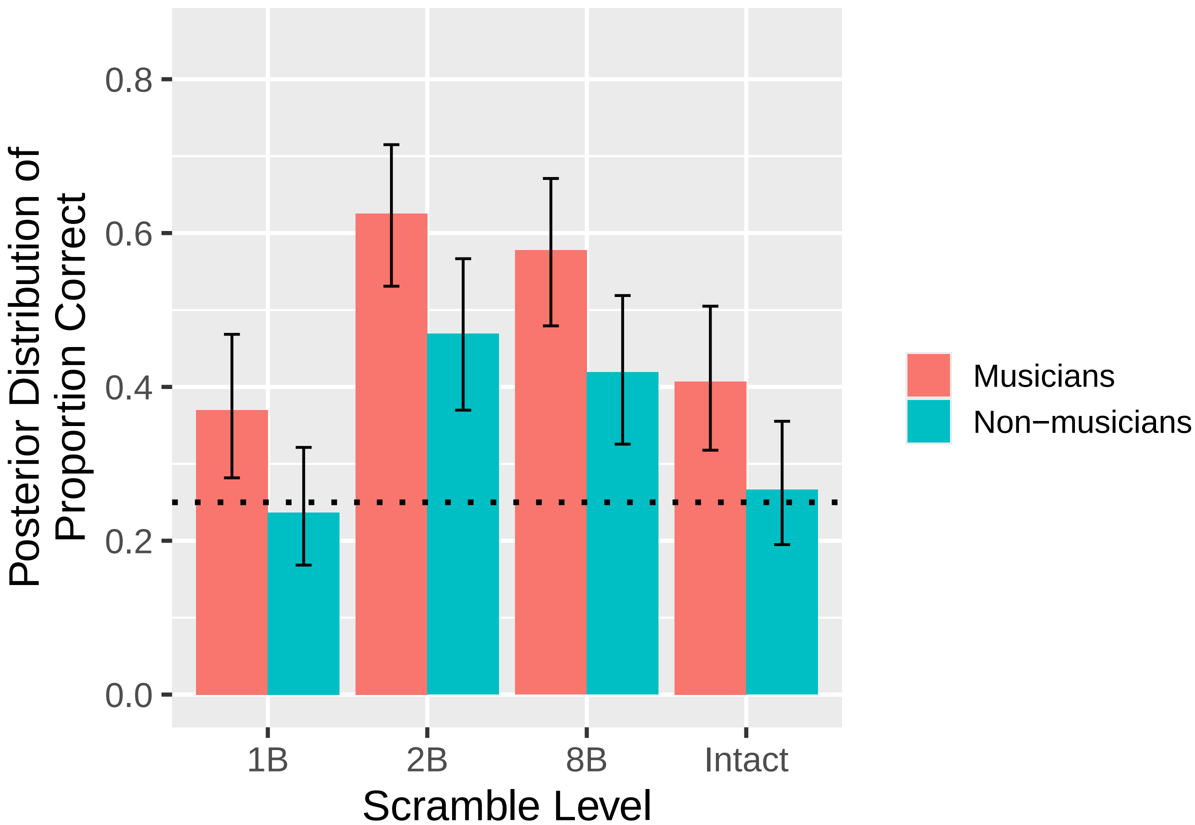
```
calculate_prob_from_logodds <- function(logodds) {
  return(exp(logodds) / (1 + exp(logodds)))
}
```

Visualize with posterior estimates and 95% CrI on the scale of accuracy.

```
posterior_est <- as.data.frame(emm_mus_scram_ms)
```

```
ggplot() +
  geom_col(aes(x = scramble, y = calculate_prob_from_logodds(emmean), fill = Musician),
           data = posterior_est,
           position = "dodge") +
  geom_errorbar(aes(x = scramble,
                    ymin = calculate_prob_from_logodds(lower.HPD),
                    ymax = calculate_prob_from_logodds(upper.HPD),
                    fill = Musician),
                data = posterior_est, position = position_dodge(width = 0.9), width = 0.2) +
  geom_hline(yintercept = 0.25, linetype = "dotted", color = "black", linewidth = 1) +
  theme_gray(base_size = 16) +
  scale_x_discrete(limits = rev) +
  ylim(0, 0.85) +
  xlab('Scramble Level') +
  ylab('Posterior Distribution of\nProportion Correct') +
  scale_fill_discrete(name="", labels=c('Musicians', 'Non-musicians')) +
  theme(legend.text = element_text(size = 12))
```

```
## Warning in geom_errorbar(aes(x = scramble, ymin =
## calculate_prob_from_logodds(lower.HPD), : Ignoring unknown aesthetics: fill
```




```r
ggsave('../figures/Fig5_categorization.png', width = 7, height = 5)
```

Years of experience

Keep only the subjects for which we have years of experience data and average accuracy per condition.

```
yrs_exp <- data %>%
  filter(!is.na(yrs_mus_exp)) %>%
  group_by(exp_subject_id, scramble, yrs_mus_exp) %>%
  summarize(count = n(),
            n_correct = sum(response),
            accuracy = n_correct / count)
```

```
## `summarise()` has grouped output by 'exp_subject_id', 'scramble'. You can
## override using the `.groups` argument.
```

Priors

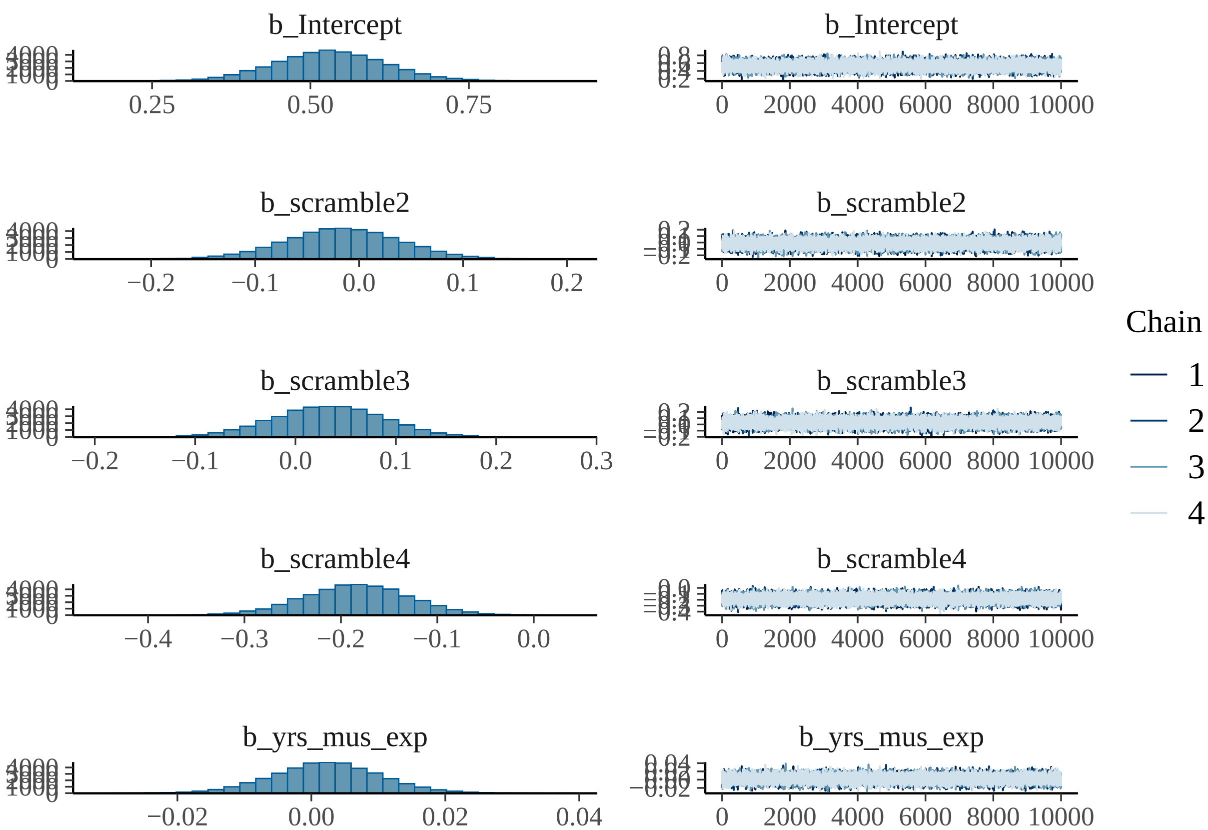
For this analysis, we're operating on the scale of accuracy. Because we don't see ceiling effects (i.e. participants aren't getting too close to perfect accuracy), a linear model is appropriate enough.

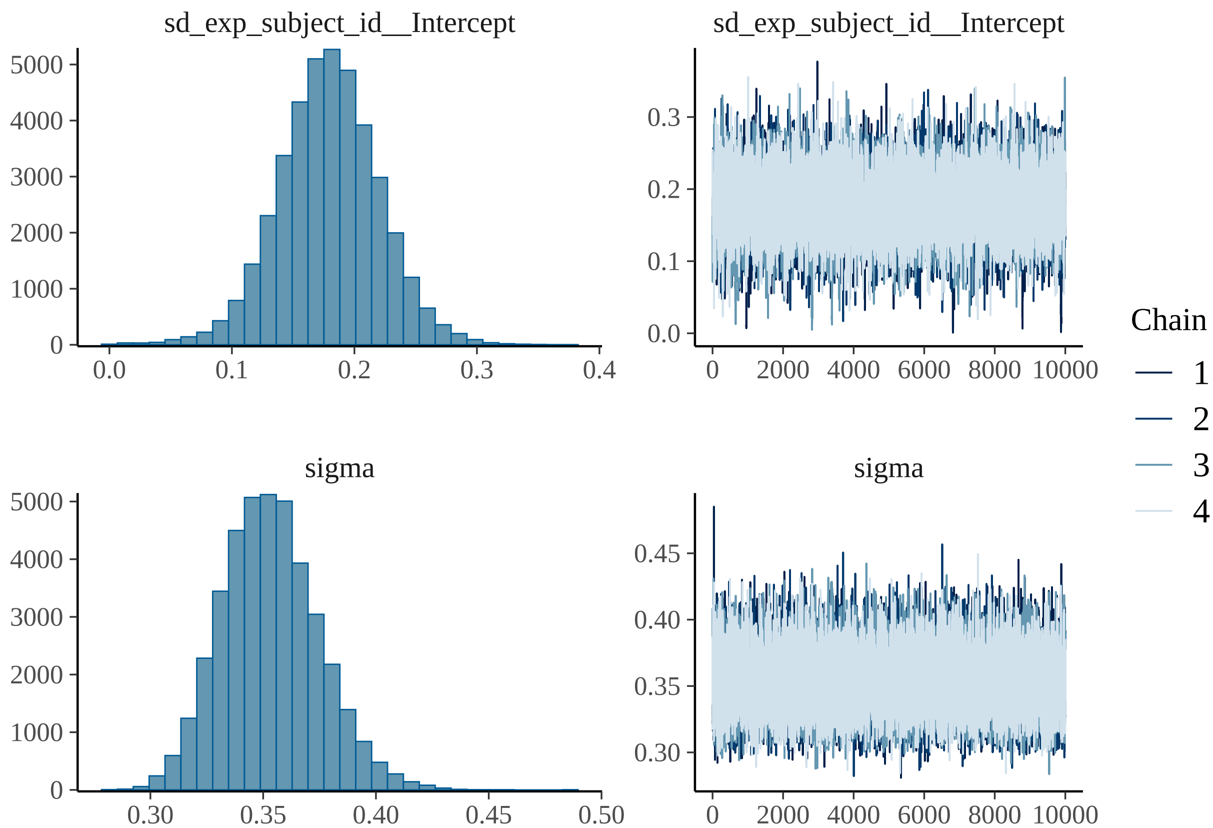
```
these_priors <- c(
  set_prior('normal(0.625, 0.1)', class = 'Intercept'),
  set_prior('normal(-0.1, 0.1)', coef = 'scramble2'),
  set_prior('normal(-0.2, 0.1)', coef = 'scramble3'),
  set_prior('normal(-0.3, 0.1)', coef = 'scramble4'),
  set_prior('normal(0, 0.1)', coef = 'yrs_mus_exp')
)
```

Main model

```
years_mus_scram <- brm(accuracy ~ scramble + yrs_mus_exp + (1|exp_subject_id), data = yrs_exp,
                       prior = these_priors,
                       save_pars = save_pars(all = TRUE), iter = 20000, refresh = 0,
                       file = 'models/E4_years')
```

```
plot(years_mus_scram)
```






```r
print(summary(years_mus_scram), digits = 4)
```

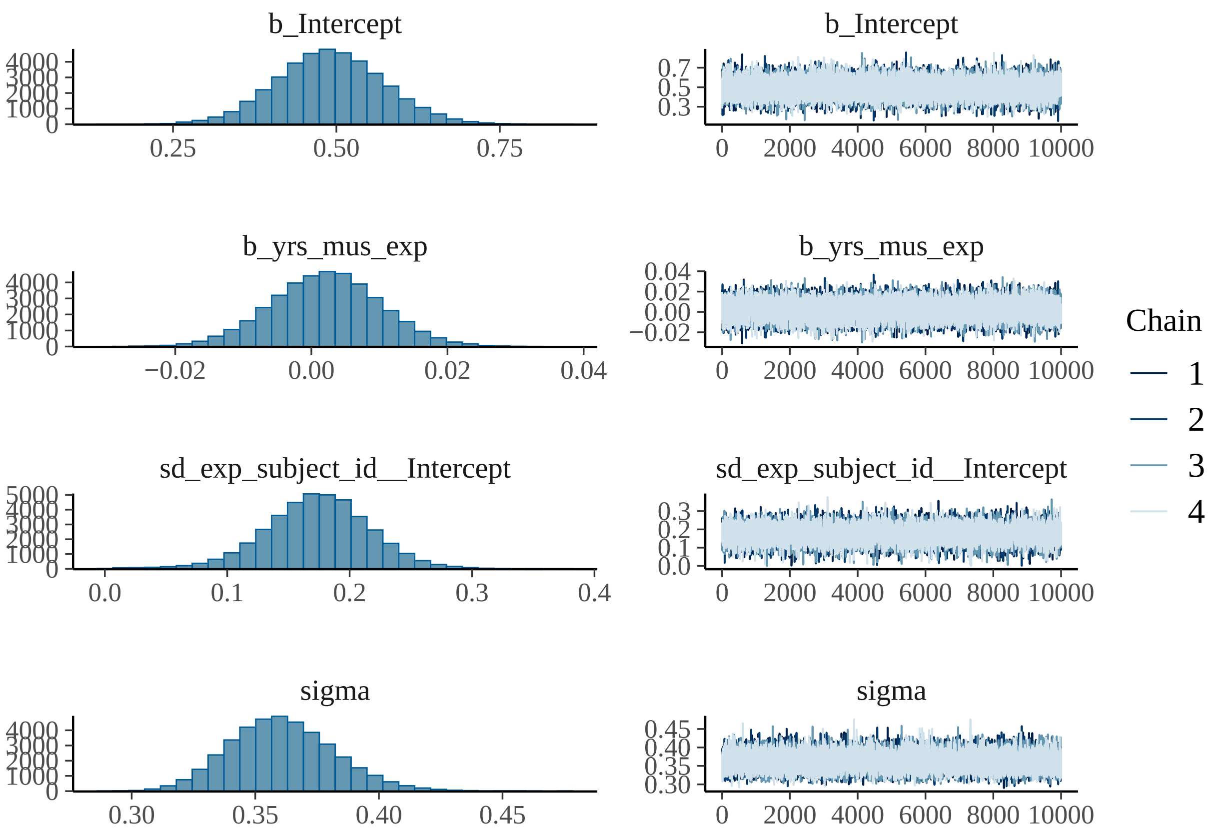
```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: accuracy ~ scramble + yrs_mus_exp + (1 | exp_subject_id)
##    Data: yrs_exp (Number of observations: 203)
##   Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
##          total post-warmup draws = 40000
##
## Multilevel Hyperparameters:
## ~exp_subject_id (Number of levels: 51)
##               Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sd(Intercept)   0.1790    0.0412   0.0970   0.2601 1.0002    11249    13637
##
## Regression Coefficients:
##             Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## Intercept     0.5312    0.0875   0.3604   0.7044 1.0000    32698    28491
## scramble2    -0.0155    0.0551  -0.1250   0.0924 1.0002    49197    31529
## scramble3     0.0328    0.0553  -0.0760   0.1394 1.0000    49052    32666
## scramble4    -0.1836    0.0550  -0.2929  -0.0765 1.0000    49417    32925
## yrs_mus_exp   0.0023    0.0080  -0.0136   0.0181 1.0000    32125    29895
##
## Further Distributional Parameters:
##       Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sigma   0.3532    0.0215   0.3142   0.3983 1.0001    24441    27563
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
```

```
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Null model (for plotting purposes)

```r
years_mus <- brm(accuracy ~ yrs_mus_exp + (1|exp_subject_id), data = yrs_exp,
                 prior = c(
                   set_prior('normal(0.625, 0.1)', class = 'Intercept'),
                   set_prior('normal(0, 0.1)', coef = 'yrs_mus_exp')),
                 save_pars = save_pars(all = TRUE), iter = 20000, refresh = 0,
                 file = 'models/E4_years_null')
```

```r
plot(years_mus)
```




```r
print(summary(years_mus), digits = 4)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: accuracy ~ yrs_mus_exp + (1 | exp_subject_id)
##    Data: yrs_exp (Number of observations: 203)
##   Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
##          total post-warmup draws = 40000
##
## Multilevel Hyperparameters:
## ~exp_subject_id (Number of levels: 51)
##                Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.1735    0.0431   0.0845   0.2556 1.0002    10058    10953
##
## Regression Coefficients:
```

```
##             Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## Intercept     0.4906    0.0821   0.3311   0.6538 1.0000    41201    30923
## yrs_mus_exp   0.0021    0.0080  -0.0138   0.0179 1.0000    40497    31150
##
## Further Distributional Parameters:
##       Estimate Est.Error l-95% CI u-95% CI   Rhat Bulk_ESS Tail_ESS
## sigma   0.3610    0.0212   0.3224   0.4052 1.0000    25917    23933
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
yrs_BF <- describe_posterior(years_mus_scram,
                             estimate = "median", dispersion = TRUE,
                             ci = .95, ci_method = "HDI",
                             test = c("bayes_factor"))
print(yrs_BF, digits = 4)
```

```
## Summary of Posterior Distribution
##
## Parameter   |  Median |    MAD |         95% CI |       BF |  Rhat |        ESS
## --------------------------------------------------------------------------------
## (Intercept) |  0.5308 | 0.0866 | [ 0.36,  0.70] | 7.77e+04 | 1.000 | 32738.0000
## scramble2   | -0.0156 | 0.0548 | [-0.12,  0.10] |    0.349 | 1.000 | 48817.0000
## scramble3   |  0.0333 | 0.0554 | [-0.08,  0.14] |    0.087 | 1.000 | 48990.0000
## scramble4   | -0.1834 | 0.0545 | [-0.29, -0.08] |     1.26 | 1.000 | 49511.0000
## yrs_mus_exp |  0.0023 | 0.0078 | [-0.01,  0.02] |    0.083 | 1.000 | 32026.0000
```

```r
yrs_null_BF <- describe_posterior(years_mus,
                                  estimate = "median", dispersion = TRUE,
                                  ci = .95, ci_method = "HDI",
                                  test = c("bayes_factor"))
print(yrs_null_BF, digits = 4)
```

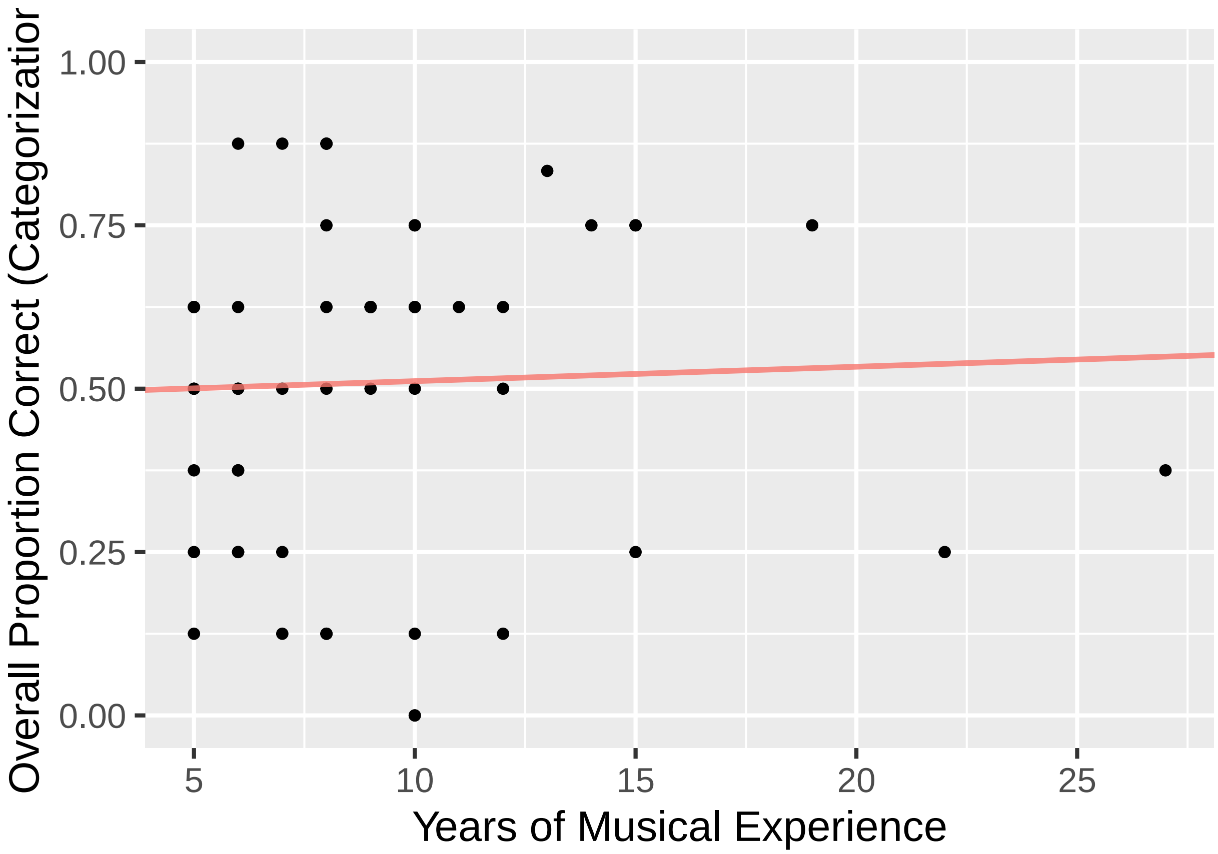
```
## Summary of Posterior Distribution
##
## Parameter   | Median |    MAD |        95% CI |       BF |  Rhat |        ESS
## ------------------------------------------------------------------------------
## (Intercept) | 0.4895 | 0.0808 | [ 0.34, 0.66] | 9.95e+03 | 1.000 | 41339.0000
## yrs_mus_exp | 0.0022 | 0.0079 | [-0.01, 0.02] |    0.083 | 1.000 | 40379.0000
```

Strong evidence against an effect of years of musical experience.

Figure S1C

```
yrs_exp %>%
  group_by(exp_subject_id, yrs_mus_exp) %>%
  summarize(mean_acc = mean(accuracy)) %>%
  ggplot(aes(yrs_mus_exp, mean_acc)) +
  geom_point() +
  geom_abline(intercept = yrs_null_BF$Median[1], slope = yrs_null_BF$Median[2],
              color = '#F8766D', linewidth = 1, alpha = 0.8) +
  xlab('Years of Musical Experience') +
  ylab('Overall Proportion Correct (Categorization)') +
  scale_x_continuous(breaks = seq(5,30,5)) +
  scale_y_continuous(breaks = seq(0, 1, 0.1)) +
  ylim(0,1) +
  theme_gray(base_size = 16)
```

```
## `summarise()` has grouped output by 'exp_subject_id'. You can override using
## the `.groups` argument.
## Scale for y is already present. Adding another scale for y, which will replace
## the existing scale.
```




```
ggsave('../figures/FigS1C_categorization.png', width = 5, height = 5)
```