# E1_report

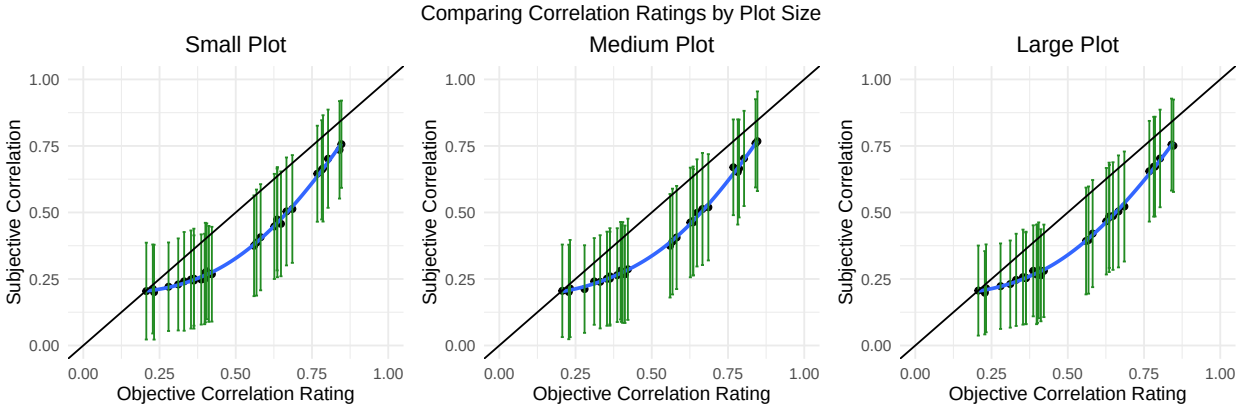
Gabriel Strain

21/03/2022

## Performance on Correlation Estimation is Better on Larger Plots. In addition, the Presence of a Novel Contrast Encoding also Improves Performance.

The plots below show subjective correlation performance for Small, Medium, and Large plots. These sizes corresponded to 63%, 100% and 252% size ratios. We hypothesised that performance at estimating correlation would be better when the plot was larger. N = 112, 180 experimental trials per participant (fully repeated measures: all participants saw all versions of each plot).

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```

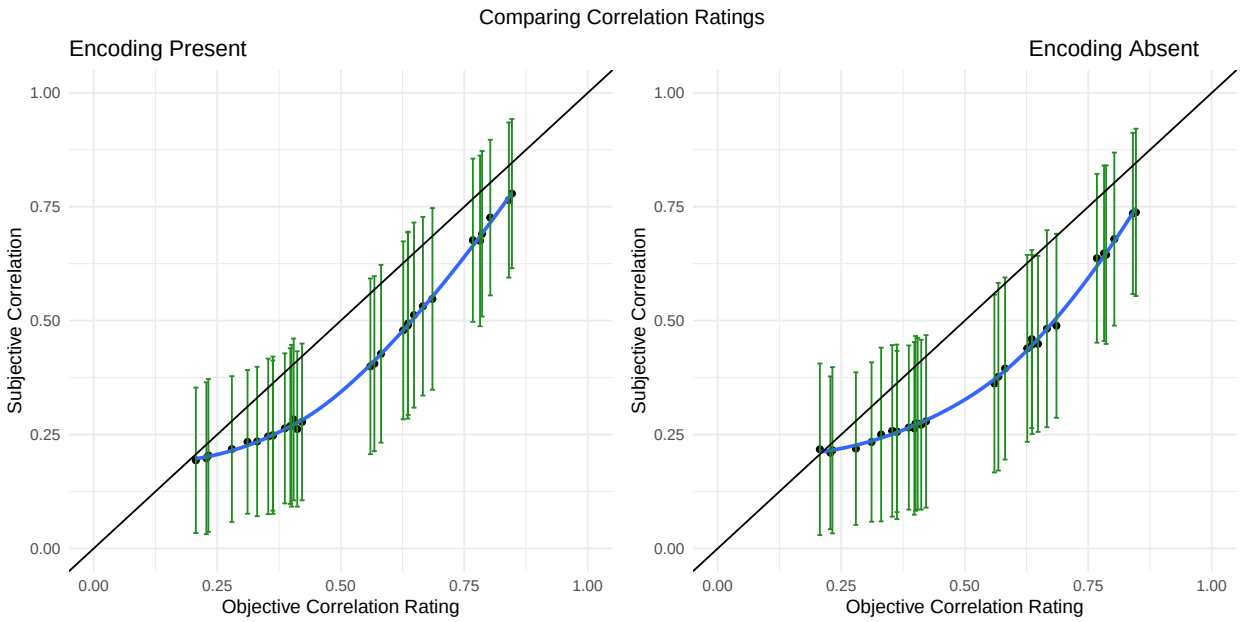



The summary statistics for the factor of size are as follows.

```
## # A tibble: 3 x 3
##   size   mean    sd
##   <fct> <dbl> <dbl>
## 1 L     0.114 0.189
## 2 M     0.116 0.192
## 3 S     0.123 0.191
```

We utilised a novel contrast encoding whereby the contrast of a certain point was linearly related to the size of the residual of that point. The presence of this contrast encoding was associated with better performance on correlation estimation. Summary statistics are below the plots.

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```




```
## # A tibble: 2 x 3
##   present  mean    sd
##   <fct>   <dbl> <dbl>
## 1 N       0.127 0.197
## 2 Y       0.109 0.183
```

# Modelling

A Linear Mixed Model was fitted with the **buildmer** package, which begins with a maximal model and drops terms until it converges and no errors are returned. The model built had the following syntax;

```
model <- lmer(difference ~ size * present + (1 | participant) + (1 | item))
```

This model has random intercepts for participants and items.

An ANOVA comparing models for participant's difference scores revealed a significant difference between estimations of correlation for difference levels of the two factors: $\chi^2(5) = 88.15, p < .001$.

There was no effect of an interaction between the two factors. Estimated marginal means and contrasts for the two factors can be seen below:

```
## $emmeans
##  size emmean    SE  df asymp.LCL asymp.UCL
##  L     0.114 0.014 Inf    0.0871     0.142
##  M     0.116 0.014 Inf    0.0889     0.144
##  S     0.123 0.014 Inf    0.0957     0.150
##
## Results are averaged over the levels of: present
```

```
## Degrees-of-freedom method: asymptotic
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate     SE  df z.ratio p.value
##  L - M    -0.00186 0.0025 Inf  -0.745  0.7369
##  L - S    -0.00861 0.0025 Inf  -3.441  0.0017
##  M - S    -0.00675 0.0025 Inf  -2.696  0.0192
##
## Results are averaged over the levels of: present
## Degrees-of-freedom method: asymptotic
## P value adjustment: tukey method for comparing a family of 3 estimates

## $emmeans
##  present emmean     SE  df asymp.LCL asymp.UCL
##  N        0.127 0.0139 Inf    0.0993     0.154
##  Y        0.109 0.0139 Inf    0.0819     0.136
##
## Results are averaged over the levels of: size
## Degrees-of-freedom method: asymptotic
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate      SE  df z.ratio p.value
##  N - Y      0.0174 0.00204 Inf   8.532  <.0001
##
## Results are averaged over the levels of: size
## Degrees-of-freedom method: asymptotic
```

## Modelling with Graph Literacy as a Fixed Effect

Participants also answered a 5-item questionnaire on Subjective Graph Literacy (Garcia-Retamero et al., 2016). We build a model with graph literacy as a fixed effect

An ANOVA between the model including literacy as a fixed effect and the original model reveals no significant effect of graph literacy: $\chi^2 = 2.01(1), p = 0.16$.

## Modelling with Residuals

As two scatterplots can show the same correlations, but have different point-by-point residuals, we can also model to include the sum of the residuals as a predictor.

An ANOVA between the model including total residuals as a fixed effect and the original model reveals no significant effect of total residual size for each plot: $\chi^2(1), p = 0.12$