March 31st
using ANOVA
Assumptions of ANOVA

1. random sampling from source population
Assumptions of ANOVA

1. random sampling from source population

2. independent measures within each sample, yielding uncorrelated response residuals
Assumptions of ANOVA

1. random sampling from source population

2. independent measures within each sample, yielding uncorrelated response residuals

3. homogenous variances across all sampled populations
Assumptions of ANOVA

1. random sampling from source population

2. independent measures within each sample, yielding uncorrelated response residuals

3. homogenous variances across all sampled populations

4. normal distribution of response residuals around the model
Assumptions of ANOVA

1. random sampling from source population

2. independent measures within each sample, yielding uncorrelated response residuals

3. homogenous variances across all sampled populations

4. normal distribution of response residuals around the model
Luka (2005) structural priming of acceptability judgments
Luka (2005) structural priming of acceptability judgments

The brewmaster who sprouted the barley selected the hops.
Luka (2005) structural priming of acceptability judgments

The brewmaster who sprouted the barley selected the hops.

The veterinarian who vaccinated the horse harnessed the dappled mule.
Luka (2005) structural priming of acceptability judgments

The brewmaster who sprouted the barley selected the hops.

The veterinarian who vaccinated the horse harnessed the dappled mule.

1  2  3  4  5  6  7

Very ungrammatical  Perfectly grammatical
Fig. 2. Mean grammaticality ratings from Experiment 4 (error bars denote SE).

Luka (2005) expt 4

Fig. 3. Mean grammaticality ratings from Experiment 5 (error bars denote SE).
grammatical sentences, particularly for structural repetition. Although there were no interactions with Grammaticality, inspection of the means prompted follow-up tests. One way ANOVAs conducted on the means ratings of sentences familiar from the reading phase versus novel confirmed di

\[ F_1(1, 25) = 10.34, p < .005; \]
\[ F_2(1, 23) = 16.31, p < .001. \]

For structural repetitions, the follow-up contrast was not significant by participants and items,

\[ F_1(1, 25) = 0.98, p = .3; \]
\[ F_2(1, 43) = 0.63, p = .4. \]

Thus, while structural facilitation was strong for moderately grammatical sentences, variability in the ratings for highly grammatical sentences reduced the overall effect of facilitation for structurally related sentences. As in Experiment 3, it seems that sentences with high initial ratings have less room for improvement due to prior exposure.

In the questionnaire following the main experiment, participants first answered the general question "Did you notice any relation between the sentences you read aloud during that tape recorded section and the sentences presented for rating?" All responded "yes." The questionnaire then included two questions about identical repetition: (a) "While you were doing the rating task, did you notice whether any of the sentences were repeated from the reading part of the task?" and (b) "You read about 50 sentences in the tape recorded part of the experiment. None, some, or all of these sentences may have appeared again for you to rate. How many of the original 50 sentences were identically repeated (word-for-word) in the rating part of the task?" Two additional questions asked about structural repetition: (c) "None, some, or all of the 50 sentences in the tape recorded task may have been similar to sentences in the rating task without being identical repetitions. When you were doing the rating task, were you aware of any type of similarity between the recorded sentences and the sentences you rated? Yes/No How would you describe the similarity?", and (d) "How many of the original 50 sentences were similar to sentences you rated (but not identically repeated)?"

The exit questionnaire indicated that nearly all of the participants were aware of the identical repetition manipulation; 22 of 24 responded "yes" when asked if they noticed sentences in the rating task that had occurred in the reading task. When asked how many of the reading-phase items re-occurred in the rating task, the mean estimate was 14.5 (SE = 1.8; the correct answer was 12). A smaller majority, 18 of 24 participants, indicated awareness of the structural repetition manipulation; the mean estimate of the number of structurally similar sentences was 11.5 (SE = 2.0; correct answer = 12). Not all of the participants chose to respond to the open-ended question asking them to describe any similarity they observed, but two of the responses included: "Grammatical errors were alike in both sections." "Wording of the sentences was in the same pattern." Other responses by participants were at least somewhat ambiguous: "Subject matter was similar." "Words switched." "Same type of words." "Fairly close." "Word type used in each." and "General trend." It is unclear to what degree participants may have been guessing. The role of conscious awareness for structural similarity remains open for investigation.

Fig. 2. Mean grammaticality ratings from Experiment 4 (error bars denote SE).
What’s the model?

Call:
\texttt{lm(formula = RATING ~ GRAMMATICALITY * FAMILIARITY * TYPE.TOKEN, }
\texttt{data = LB1)}

Residuals:
\begin{tabular}{c c c c c}
Min & 1Q & Median & 3Q & Max \\
-2.7026 & -0.3673 & 0.1269 & 0.4936 & 1.9487
\end{tabular}
What's the model?

Call:
lm(formula = RATING ~ GRAMMATICALITY * FAMILIARITY * TYPE.TOKEN, 
    data = LB1)

Residuals:
    Min     1Q Median     3Q    Max
-2.7026 -0.3673  0.1269  0.4936  1.9487
Response residuals are not normally-distributed

```
qqnorm(residuals(fitted))
```
Transform the data

### TABLE 3.8
THE "LADDER OF POWERS" FOR TRANSFORMING SCORES WHERE $Y' = Y^p$

<table>
<thead>
<tr>
<th>Power, $p$</th>
<th>Transformation</th>
<th>Name and Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$Y^2$</td>
<td>Square, useful when need to increase spread among higher scores relative to lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scores, e.g. with negatively skewed data</td>
</tr>
<tr>
<td>1</td>
<td>$Y$</td>
<td>Raw data, no transformation</td>
</tr>
<tr>
<td>$1/2$</td>
<td>$\sqrt{Y}$</td>
<td>Square root, helpful with positively skewed distributions</td>
</tr>
<tr>
<td>0</td>
<td>$\log_{10} Y$</td>
<td>Logarithmic, use when $p$ estimated to be near zero, e.g., with extremely positively</td>
</tr>
<tr>
<td></td>
<td></td>
<td>skewed distributions. Use $Y' = \log_{10}(Y + 1)$ if there are any zeros in data.</td>
</tr>
<tr>
<td>-1</td>
<td>$1/Y$</td>
<td>Reciprocal, e.g., transforming latency to speed. Use $Y' = 1/(Y + 1)$ if there are</td>
</tr>
<tr>
<td></td>
<td></td>
<td>any zeros in data.</td>
</tr>
</tbody>
</table>
ANOVA on arcsine-transformed ratings
ANOVA on arcsine-transformed ratings
ANOVA on arcsine-transformed ratings
Arcsine of the square root is positive
Arcsine of the square root is positive

Scale to \([0,1]\) by multiplying by reciprocal of \(\pi/2\)
Model of arcsine-transformed ratings

\[ y = \frac{2}{\pi} \sin^{-1}(\sqrt{x}) \]

\[
\text{lm(formula = } 2/\pi \times \text{asin(sqrt((RATING/7))) } ~ \text{GRAMMATICALITY } \times \\
\text{ FAMILIARIETY } \times \text{ TYPE.TOKEN, data = LB1})
\]

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.25962</td>
<td>-0.05240</td>
<td>0.00566</td>
<td>0.05962</td>
<td>0.34377</td>
</tr>
</tbody>
</table>
Model of arcsine-transformed ratings

\[ y = \frac{2}{\pi} \sin^{-1}(\sqrt{x}) \]

\[
\text{lm(formula = } \frac{2}{\pi} \text{ * asin(sqrt((RATING/7))) } ~ \text{ GRAMMATICALITY } * \text{ FAMILIARITY } * \text{ TYPE.TOKEN, data = LB1)}
\]

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>-0.25962</td>
<td>-0.05240</td>
<td>0.00566</td>
<td>0.05962</td>
<td>0.34377</td>
</tr>
</tbody>
</table>
Helped

Normal Q-Q Plot

Sample Quantiles

Theoretical Quantiles

outlier
Random effects

Expected Mean-Squares for Repeated-Measurements ANOVA with Words Sampled Randomly

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Label</th>
<th>df</th>
<th>Expected mean-squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>A</td>
<td>$p - 1$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + q\sigma_{AS}^2 + n\sigma_{W(A)}^2 + nq\sigma_A^2$</td>
</tr>
<tr>
<td>Words (within Treatment)</td>
<td>W(A)</td>
<td>$p(q - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + n\sigma_{W(A)}^2$</td>
</tr>
<tr>
<td>Subjects</td>
<td>S</td>
<td>$n - 1$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + pq\sigma_S^2$</td>
</tr>
<tr>
<td>Treatment × Subjects</td>
<td>AS</td>
<td>$(p - 1)(n - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + q\sigma_{AS}^2$</td>
</tr>
<tr>
<td>Words × Subjects</td>
<td>W(A)S</td>
<td>$p(q - 1)(n - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2$</td>
</tr>
</tbody>
</table>

Note. $p =$ number of levels of the treatment variable; $n =$ number of subjects; $q =$ number of items. Words and Subjects are assumed to be random effects.
Usual way of testing for an effect

\[ F = \frac{E(MS_b) = \sigma^2_{\epsilon} + n\sigma^2_\alpha}{E(MS_w) = \sigma^2_{\epsilon}} \]
Usual way of testing for an effect

\[ F = \frac{E (M S_b)}{E (M S_w)} = \frac{\sigma^2_\varepsilon + n\sigma^2_\alpha}{\sigma^2_\varepsilon} \]

pushes \( F > 1 \) if there is a treatment effect.
**Quasi F-ratio**

**Expected Mean-Squares for Repeated-Measurements ANOVA with Words Sampled Randomly**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Label</th>
<th>df</th>
<th>Expected mean-squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>A</td>
<td>$p - 1$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + q\sigma_{AS}^2 + n\sigma_{W(A)}^2 + nq\sigma_A^2$</td>
</tr>
<tr>
<td>Words (within Treatment)</td>
<td>W(A)</td>
<td>$p(q - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + n\sigma_{W(A)}^2$</td>
</tr>
<tr>
<td>Subjects</td>
<td>S</td>
<td>$n - 1$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + pq\sigma_S^2$</td>
</tr>
<tr>
<td>Treatment × Subjects</td>
<td>AS</td>
<td>$(p - 1)(n - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2 + q\sigma_{AS}^2$</td>
</tr>
<tr>
<td>Words × Subjects</td>
<td>W(A)S</td>
<td>$p(q - 1)(n - 1)$</td>
<td>$\sigma_e^2 + \sigma_{W(A)S}^2$</td>
</tr>
</tbody>
</table>

*Note.* $p =$ number of levels of the treatment variable; $n =$ number of subjects; $q =$ number of items. Words and Subjects are assumed to be random effects.

\[
F' = \frac{MS_A + MS_{W(A)S}}{MS_{AS} + MS_{W(A)}}
\]
Expected Mean-Squares for Repeated-Measurements ANOVA with Words Sampled Randomly

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Label</th>
<th>df</th>
<th>Expected mean-squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>A</td>
<td>$p - 1$</td>
<td>$\sigma^2_e + \sigma^2_{W(A)S} + q\sigma^2_{AS} + n\sigma^2_{W(A)} + nq\sigma^2_A$</td>
</tr>
<tr>
<td>Words (within Treatment)</td>
<td>W(A)</td>
<td>$p(q - 1)$</td>
<td>$\sigma^2_e + \sigma^2_{W(A)S} + n\sigma^2_{W(A)}$</td>
</tr>
<tr>
<td>Subjects</td>
<td>S</td>
<td>$n - 1$</td>
<td>$\sigma^2_e + \sigma^2_{W(A)S} + pq\sigma^2_S$</td>
</tr>
<tr>
<td>Treatment × Subjects</td>
<td>AS</td>
<td>$(p - 1)(n - 1)$</td>
<td>$\sigma^2_e + \sigma^2_{W(A)S} + q\sigma^2_{AS}$</td>
</tr>
<tr>
<td>Words × Subjects</td>
<td>W(A)S</td>
<td>$p(q - 1)(n - 1)$</td>
<td>$\sigma^2_e + \sigma^2_{W(A)S}$</td>
</tr>
</tbody>
</table>

*Note. $p = \text{number of levels of the treatment variable}; n = \text{number of subjects}; q = \text{number of items}. Words and Subjects are assumed to be random effects.*

$$F' = \frac{MS_A + MS_{W(A)S}}{MS_{AS} + MS_{W(A)}}$$

not based on independent sum-of-squares; only approximates an F distribution
A lower bound on the quasi F-ratio

\[ \min F' = \frac{MS_A}{MS_{AS} + MS_{W(A)}} = \frac{F_1 F_2}{F_1 + F_2} \]
A lower bound on the quasi F-ratio

Table 4.9 MinF’ values (and the error degrees of freedom) in the analysis of Luka and Barsalou’s data. The F-values found in the subjects analysis are shown in the first column. The second column shows the F-values from the items analysis. F-values that are significant at p < 0.05 are underlined.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>minF’</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>grammaticality</td>
<td>221.0</td>
<td>77.0</td>
<td>57.10</td>
<td>65</td>
</tr>
<tr>
<td>familiarity</td>
<td>10.5</td>
<td>9.4</td>
<td>4.96</td>
<td>65</td>
</tr>
<tr>
<td>repetition type</td>
<td>5.6</td>
<td>1.7</td>
<td>1.32</td>
<td>63</td>
</tr>
<tr>
<td>fam*rep</td>
<td>1.4</td>
<td>3.5</td>
<td>1.00</td>
<td>45</td>
</tr>
<tr>
<td>gram<em>fam</em>rep</td>
<td>3.7</td>
<td>1.3</td>
<td>0.95</td>
<td>65</td>
</tr>
</tbody>
</table>
Question 4 on page 141 of Johnson