1. (d) both b and c

[Condition means differ in part because measurement error causes a subject's score to vary from occasion to occasion (condition to condition), even if there is no influence of the independent variable. Condition means will vary even more if there is an effect of the independent variable on scores (also known as treatment effects). Differences between subjects will not affect condition means because the same subjects are tested in all conditions.]

2. (c) measurement error

[In all ANOVA designs, it is assumed that if there is an effect of the independent variable, it will affect all subjects equally so only the population means are influenced, whereas the population variance is unchanged. In a repeated-measures design, if subjects differ in how their scores change across conditions, this can be due only to measurement error because when the null hypothesis is correct, each subject has the same true score in each condition. When the null hypothesis is false and population means differ, it is assumed that the size of the effect of the independent variable is the same for each subject so as far as the influence of the independent variable is concerned. So each subject should show the same degree of difference between conditions.]

3. (a) variability among condition means is computed in the same way as the one-factor independent-groups ANOVA

[The formula for $SS_{conditions}$ in the repeated-measures design and the formula for $SS_{groups}$ in the one-factor, independent-sample design are the same. Differences between subject means do not affect $SS_{conditions}$ because the same subjects are tested in each condition. Total variability in the repeated-measures design is computed by taking deviations between raw scores and the grand mean, as in any ANOVA design.]

4. (a) 16  \[df_{subjects} = 16 - 1 = 15\] \hspace{1cm} (b) 4  \[df_{conditions} = 4 - 1 = 3\]

5. (a) Possible set of condition orders using a latin-square counterbalancing: ABC, BCA, CAB.

(b) Two subjects would be tested in each order.

(c) ANOVA summary table:
6. The condition means are as follows: Cong. = 714, Incong. = 782, and Neutral = 720. For claim 1, the comparison is Incong. vs. Neutral. If incongruence interferes with color naming, then naming should take longer in the Incong. condition than in the Neutral condition. Comparing Incong. and Cong. conditions would not provide a clear answer to the question about interference because a difference between those two conditions could be a result of interference due to incongruence, benefit due to congruence, or a combination of the two. A pure measure of interference can be obtained by comparing Incong. and Neutral conditions:

\[ t = \frac{\bar{X}_{\text{Incong.}} - \bar{X}_{\text{Neutral}}}{\sqrt{\frac{s^2_{\text{Incong.}}}{n_{\text{Incong.}}} + \frac{s^2_{\text{Neutral}}}{n_{\text{Neutral}}}}} \]

This \( t \) ratio is significant [with \( df = 8 \), critical \( t = \pm 2.306 \)], so the researcher can conclude that using an incongruent word slows color naming, relative to a neutral condition.

For claim 2, the comparison is Cong. vs. Neutral (for the reason discussed above):

\[ t = \frac{\bar{X}_{\text{Cong.}} - \bar{X}_{\text{Neutral}}}{\sqrt{\frac{s^2_{\text{Cong.}}}{n_{\text{Cong.}}} + \frac{s^2_{\text{Neutral}}}{n_{\text{Neutral}}}}} \]

This \( t \) ratio is not significant, so the researcher can conclude that there is no evidence from this study that using a congruent word will improve color naming, relative to a neutral condition. The power of this test (a related-samples \( t \) test) to detect a medium effect (.5 standard deviation) depends in part on the correlation between scores in the Cong. and Neutral conditions computed across subjects. That correlation is \( r = .95 \) in this data set. This correlation can be computed using the formula for \( r \) learned in PSYC 300A, or by using SPSS. Students will not be asked to compute \( r \) on an examination in this course, but are expected to compute it for this problem so that they understand the role that \( r \) plays in the estimate of power. Note that this sample-based value of \( r \) is used as an estimate for \( \rho \) in the formula for power. Thus,
so from Table E.5 in Howell's book, power is between .94 and .95.

7. In their current arrangement, the data produce an $F$ ratio of 1.26. In general, to produce a larger $F$ ratio, the scores should be rearranged to create a more consistent pattern of condition effects across subjects. Notice that scores generally are larger in condition B than in A or C, which are similar. The scores in condition A can be rearranged so that the highest A score is paired with the highest B score, and so on. The same can be done in condition C. The result would be the following (consider constructing a line graph plotting these data and another plotting the original pattern of data):

<table>
<thead>
<tr>
<th>Subj.</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

These data will yield $F = 6.50$, which is significant with $\alpha = .05$.

To produce an $F$ ratio that is smaller than that for the original data, the consistency of the pattern of effects (B scores higher than A or C) must be reduced. That is, scores in A should be rearranged so that some subjects have a higher score in A than in B and the reverse should be true for other subjects. Inconsistency will be especially high if there are subjects who have patterns of effects that go strongly in opposite directions (compare subjects 1 and 4 below). The same can be done with the scores in condition C. One possible result is the following (plot these data as well to see how strong the inconsistency appears to be):

<table>
<thead>
<tr>
<th>Subj.</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

These data will yield $F = 0.78$.

8. Each subject is tested in each condition, so all condition means are based on the same subjects. Differences between means arise only from effects of the independent variable on an individual’s behavior, or from measurement error.

9. (a) 4 conditions, 9 subjects
   (b) critical $F(3, 24) = 3.01$
   (c) Reject $H_0$
10. (a) baseline vs. treatment; indicates whether behavior problems are reduced by the use of a token economy
    (b) treatment vs. withdrawal; indicates whether behavior problems return after the token economy is stopped
    (c) baseline vs. withdrawal; indicates whether behavior after the token economy is stopped returns to baseline, gets worse than baseline, or is an improvement over baseline

11. (a) \( H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \) \( H_1 : \) not all population means are equal
    (b) To prevent confounding of source and saying. If the same attribution assignments were used for all subjects and if differences between sources were found, then it would not be possible to determine whether those differences were due to the effect of source or to differences between sayings.
    
    (c) Source | SS  | df | MS  | F  |
    Subjects  | 57.71 | 5  | 11.54 |   |
    Attribution | 61.79 | 3  | 20.60 | 3.94 |
    Error     | 78.46 | 15 | 5.23  |   |
    Total     | 197.96 |   |       |   |
    
    critical \( F(3, 15) = 3.29 \), so the null hypothesis is rejected, \( F(3, 15) = 3.94 \)

(d) the \( t \) ratios for these three pairwise comparisons are:
    movie star vs. scientist: \((3.83-8)/\sqrt{(5.23/6 + 5.23/6)} = -3.16\)
    movie star vs. politician: \((3.83-7)/\sqrt{(5.23/6 + 5.23/6)} = -2.40\)
    movie star vs. writer: \((3.83-7.33)/\sqrt{(5.23/6 + 5.23/6)} = -2.65\)
    – all three comparisons are significant with \( \alpha = .05; \) critical \( t(15) = \pm 2.132 \)

(e) The pairwise comparisons indicate that sayings attributed to movie stars are seen as having less impact than sayings generated by other professionals.

12. Read in the data file and look at the first few lines, including the header.

    > phon = read.table(file.choose(new=F),header=T)
    > head(phon)
    subj cond  RT
    1  s01  neut  371
    2  s01  pred  396
    3  s01 unexp  448
    4  s02  neut  392
    5  s02  pred  402
    6  s02 unexp  393

    Load the psych library so the describeBy function can be used to get the descriptive data.

    > library(psych)
    > describeBy(phon$RT, phon$cond)
Run the `aov` command to compute the ANOVA and show a summary of the output:

```r
> output = aov(RT ~ cond + Error(subj/cond), phon)
> summary(output)
```

```
Error: subj
   Df Sum Sq Mean Sq F value Pr(>F)
Residuals 29 208946    7205

Error: subj:cond
     Df Sum Sq Mean Sq   F value Pr(>F)
data$cond  2  19777 9889.5  23.35 3.63e-08 ***
Residuals 58  24558     423
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

There clearly is a significant effect, so use the `pairwise.t.test` function to find where the differences are.

```r
> pairwise.t.test(phon$RT, phon$cond, p.adjust = "none", paired = T)
```

```
Pairwise comparisons using paired t tests
data:  phon$RT and phon$cond
         neut    pred
pred  0.19  -
unexp 2.4e-05 2.9e-07
```

The tests show that in these data there is no evidence for an advantage of a highly predictable word over a neutral word ($p = .19$), but there is a clear disadvantage if the word carrying the phoneme is unexpected ($p < .001$ when compared to neutral and $p < .001$ when compared to predictable).