

Follow Up to One-way ANOVAs

Contents

1	Experimentwise Error Rate	2
1.1	Design Example 1 for Last Class	3
1.1.1	Run ANOVA	3
1.2	Which groups are different?	3
1.3	Philosophy I (Familywise) Correction for Error Rate	3
1.4	Approaches to Follow up	4
1.4.1	Confirmatory: Hypothesis Driven follow-up	4
1.4.2	Exploratory: Test everything approach	4
1.4.3	Hybrid approach	4
2	Methods of Pairwise Follow up	4
2.1	Fisher's Protected t-test (aka LSD)	4
2.1.1	Doing LSD in R	5
2.1.2	VS. independent t-tests	5
2.1.3	Effect-sizes	6
2.2	Tukey's HSD	7
2.2.1	Doing HSD in R	7
2.3	Bonferroni Family	7
2.3.1	Bonferroni Correction	8
2.3.1.1	Doing Bonferroni in R	8
2.3.2	Why is <i>a priori</i> Power affected?	8
2.3.3	Sidak method	10
2.3.3.1	Doing Sidak in R	10
2.3.4	Notes	10
3	Method of Contrast	10
3.1	Linear Contrast	10
3.1.1	Step 1	11
3.1.2	Step 2	11
3.1.3	Step 3	11
3.1.4	Step 4	11
3.1.5	Step 5	11
3.2	Doing Constrasts in R	12
3.3	Doing Multiple Contrasts in R	12
4	Suggestions on Review Process	13
5	APA Style Report	13
6	References	14

1 Experimentwise Error Rate

Too many t-tests inflates our type I error rate *per comparison* (α_{pc})

$$\alpha_{EW} = 1 - (1 - \alpha_{pc})^j$$

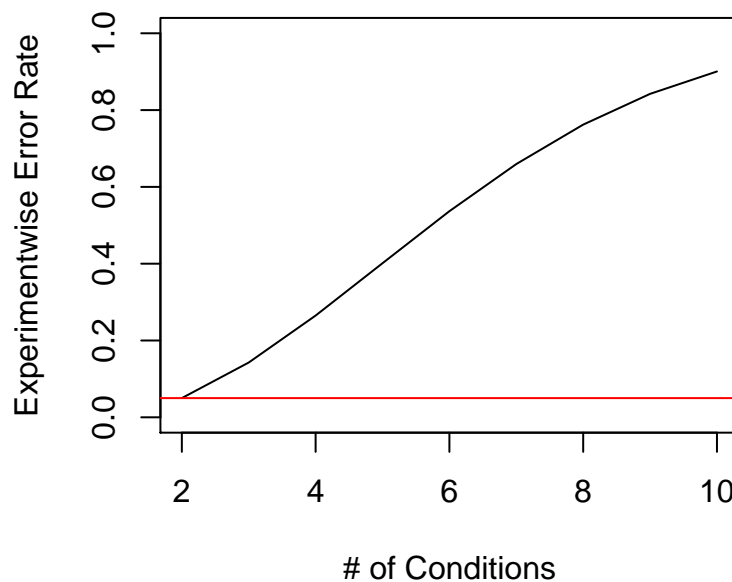
where, j = # of comparisons

$$j = \frac{K(K-1)}{2}$$

[Assuming conditional probability, each comparison is independent, and sampling with replacement]

We call this Experimentwise Error Rate (α_{EW}). If we set $\alpha_{pc} = .05$ and K = increasing from 2 to 10 we can see what happens when we have K number of conditions and we want to do all pairwise comparisons (all cells to all cells)

```
Aew <-function(alpha,K){1-(1-alpha)^(K*(K-1)/2)}  
K=seq(2,10)  
Aew.Report<-mapply(Aew, alpha=.05, K)  
  
plot(K,Aew.Report, type='l',  
      xlab = "# of Conditions", ylab="Experimentwise Error Rate", ylim=c(0,1))  
abline(h=.05, col='red')
```



Yikes!

[Note: Experimentwise is also called Familywise error]

1.1 Design Example 1 for Last Class

Interested in how ‘trusting’ of their partner people feel after engaging in various levels of joint action. 2 People enter the lab (one is a confederate) and are asked to either a) move chairs independently, b) Move a couch together, and confederate helps c) Move a couch together and confederate hinders. Moving a couch requires the two people to coordinate their action. After the study ends, people rate how trustworthy their partner seems (1-7, with 7 = high trust).

No Joint Action	Helpful Joint Action	Disrupted Joint Action
2	5	2
3	4	3
5	7	3
3	4	1
2	5	1
M = 3	M = 5	M = 2

1.1.1 Run ANOVA

```
library(afex)

ANOVA.Table<-aov_car(DV~Condition + Error(SubNum),
                     data=JA.Study)

ANOVA.Table

## Anova Table (Type 3 tests)
##
## Response: DV
##      Effect    df  MSE      F ges p.value
## 1 Condition 2, 12 1.33 8.75 ** .59    .005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

So, we need to track down which groups are different.

1.2 Which groups are different?

If we run 3 t-tests at an $\alpha_{pc} = .05$, we risk ($J=3$), $\alpha_{EW} = 0.143$.

1.3 Philosophy I (Familywise) Correction for Error Rate

Dominant Philosophy in Experimental Psychology (Not Neuroscience; We do that next week)

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	Type 1 (V)	True Positives (S)	R
Non-Sig	True Negative (U)	Type II (T)	m-R
Total	Mo	M-Mo	

V / Mo = Probability of getting at least one wrong: we want to keep this at 5%

Different correction methods proposed over the years: From the most anti-conservative to most

conservative. They come in and out of favor based on the fashion of the times within each subfield.

1.4 Approaches to Follow up

1.4.1 Confirmatory: Hypothesis Driven follow-up

Preferred approach: Look at only those conditions that you predicted would be different. That means you must carefully formulate your predictions. This is hard and will take a long time to learn. It causes people the most frustration. It is required for pre-registered reports and is being required more and more by reviewers. We are starting to leave the dark ages behind where reviewers would ask for everything to be compared to everything (because type I error problems are being better understood).

- Pros: You will do fewer tests and commit less type 1 errors. Since you run fewer tests, your follow-ups could be more anti-conservative so less type II errors as well (as you will see below).
- Cons: You might miss something interesting (but a lot of people argue that those interesting things you find are what ended up causing the replication crisis)

Summary: Low Type I & Low Type II, but risk missing something novel (assuming it's not type I)

1.4.2 Exploratory: Test everything approach

Falling out of Favor: Tests all comparisons and/or anything you find interesting. This is the easiest to understand and was a common approach, but is getting hard to publish. Requires very conservative follow up approaches because the risk of Type I is very high.

- Pros: You see everything
- Cons: If you don't do it correctly most of what you find will be Type I. If you do it correctly you will commit lots of Type II errors.

Summary: If you do it correctly, Low Type I and super high Type II. If you do it incorrectly, super high Type I and low Type II.

1.4.3 Hybrid approach

Happy Medium?: Do the Confirmatory approach (planned comparisons). Test anything exploratory as conservatively as you can (unplanned comparisons). You need to carefully distinguish what you are doing to your reader otherwise they will be confused if they are not too statistically savvy as it will seem you are choosing different corrections for unknown reasons.

2 Methods of Pairwise Follow up

2.1 Fisher's Protected t-test (aka LSD)

LSD = Least Significant Difference is the most anti-conservative as it provides *almost* no protection on type I error (but its used as the base pairwise method for other approaches). It corrects the t-test formula directly:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$LSD = \frac{M_1 - M_2}{\sqrt{\frac{MS_W}{n_1} + \frac{MS_W}{n_2}}}$$

This works because MS_W is basically S_p^2 , in that MS_W is the pooled variance term (but across all groups), where S_p^2 is just between the groups of interest. This is because the error term contains the pooled variance of all group, so its safer version of the t-test formula for our follow up tests.

LSD is only appropriate when you are doing Confirmatory testing with 3 or less comparisons

2.1.1 Doing LSD in R

LSD can be done automatically in R carrying over the error term from the ANOVA using the **emmeans** package and examining the Estimated Marginal Means (When we get to two way ANOVAs you will understand what we mean by Marginal Means). We need first to run our ANOVA and bring down the ANOVA object from **afex**. Estimated marginal means are based on a model (thus we will examine the model's estimates of the fitted values; not the raw data). [This default method for SAS and you can force SPSS with syntax to do it as well. Remember, ANOVA is a special case of regression, so this method works for regression analysis as well.]

```
library(emmeans)
Fitted.Model<-emmeans(ANOVA.Table, ~Condition)
Fitted.Model

## Condition emmean      SE df lower.CL upper.CL
## Distrupted      2 0.5163978 12 0.8748659 3.125134
## Helpful          5 0.5163978 12 3.8748659 6.125134
## No              3 0.5163978 12 1.8748659 4.125134
##
## Confidence level used: 0.95
```

You will notice we created a new object that have emmeans (means per cell), the SE (which is the same for all cells because its created from MS_W , *when sample sizes are equal* $SE = \frac{MS_w}{\sqrt{n}}$), df_w term, and CIs.

Next, we need to tell it to run LSD (which can be done multiple ways, but here is a simple way)

```
pairs(Fitted.Model, adjust='none')

## contrast      estimate      SE df t.ratio p.value
## Distrupted - Helpful      -3 0.7302967 12  -4.108  0.0015
## Distrupted - No          -1 0.7302967 12  -1.369  0.1960
## Helpful - No              2 0.7302967 12   2.739  0.0180
```

The estimate is now the mean difference between groups, SE is the formula from the LSD equation and it used df_w to look up pvalues.

2.1.2 VS. independent t-tests

- To test this I must filter `%>% droplevels` the condition I don't want so I can compare 2 groups a time. I will use the **broom** package (tidy function) so I can make my t-tests easier to view

Disrupted vs Helpful

```
library(dplyr)
library(broom)
JA.Study.1 <- JA.Study %>% filter(Condition== "Distrupted" |
                                Condition== "Helpful" ) %>%
```

```
droplevels() %>%
do(tidy(t.test(DV~Condition,
               var.equal = TRUE,
               data=)))
knitr::kable(JA.Study.1[3:5], digits=4)
```

statistic	p.value	parameter
-4.2426	0.0028	8

Disrupted vs No

```
JA.Study.2 <- JA.Study %>% filter(Condition== "Disrupted" |
                                Condition== "No" ) %>%

droplevels() %>%
do(tidy(t.test(DV~Condition,
               var.equal = TRUE,
               data=)))
knitr::kable(JA.Study.2[3:5], digits=4)
```

statistic	p.value	parameter
-1.4142	0.195	8

Helpful vs No

```
JA.Study.3 <- JA.Study %>% filter(Condition== "Helpful" |
                                Condition== "No" ) %>%

droplevels() %>%
do(tidy(t.test(DV~Condition,
               var.equal = TRUE,
               data=)))

knitr::kable(JA.Study.3[3:5], digits=4)
```

statistic	p.value	parameter
2.582	0.0325	8

You will notice the patterns are the same and that is because the variance between the cells is mostly homogeneous. LSD will be more conservative when HOV is violated (because the MS_W will be bigger than SEM for only the conditions you are comparing; Hence you have some protection).

2.1.3 Effect-sizes

- You can also get your effect sizes, but you will have to do them one at a time.

```
library(effsize)
JA.Study.E1 <- JA.Study %>% filter(Condition== "Disrupted" |
                                Condition== "Helpful" ) %>%

droplevels()
# Run Effect size
```

```
cohen.d(DV=Condition, data=JA.Study.E1)
```

```
##
## Cohen's d
##
## d estimate: 2.683282 (large)
## 95 percent confidence interval:
##      inf      sup
## 0.6729538 4.6936094
```

2.2 Tukey's HSD

John Tukey created the Honestly Significant Difference test (also called a Tukey test or Tukey Range test) to protect type I error better. This is a popular method for pairwise comparisons because its a little more conservative than LSD (but mostly still on the anti-conservative side), but does not kill your power and its easy to understand. I think people like it because they don't "lose" all their effects (but again it's not very conservative overall). The formulas are very similar to LSD, but Tukey created a new F-table, called the q-table that corrects the critical values to make them more conservative than F-table. So you would compare your q_{exp} to the q_{crit} value like a t-test.

$$q = \frac{M_1 - M_2}{\sqrt{\frac{MS_W}{n}}}$$

Here I show you the simple formula we use to calculate it by hand. It will provide the HSD value. If Mean differences > HSD value, then your mean differences are significant.

$$HSD = q_{crit} \sqrt{\frac{MS_W}{n}}$$

2.2.1 Doing HSD in R

emmeans package will provide you a t-value and corrected the p-values automatically. [Note the computer can correct for unequal sample size (Tukey-Kramer method) as the formulas I gave you do not].

```
# Fitted.Model<-emmeans(ANOVA.Table, ~Condition) #we already created this
pairs(Fitted.Model, adjust='tukey')
```

```
## contrast      estimate      SE df t.ratio p.value
## Distrupted - Helpful      -3 0.7302967 12  -4.108  0.0038
## Distrupted - No          -1 0.7302967 12  -1.369  0.3866
## Helpful - No              2 0.7302967 12   2.739  0.0441
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Note the message at the bottom (*P value adjustment: Tukey method for comparing a family of 3 estimates*). This corrected for 3 tests. We will need to be careful about this in the future when we move to 2-way ANOVAs. Also, **read the notes carefully as sometimes emmeans does not do what you request because it thinks what you are asking for is not sensical statistically.**

2.3 Bonferroni Family

The logic of these tests is to lower you α_{pc} based on the number of tests you need to run until the rate at which you commit a $\alpha_{EW} = .05$ (or the rate you set). These tests are extremely conservative and kill your power. While

they are used often by psychologists, they are not liked by methodologists (but you will be asked to do this at some point). Modern methodologists are pushing other methods that help protect Type I without inflating Type II as much.

2.3.1 Bonferroni Correction

This is the most conservative and least powerful

$$\alpha_{pc} = \frac{\alpha_{EW}}{j}$$

For our study that means:

$$\alpha_{pc} = \frac{\alpha_{EW}}{j} = \frac{.05}{3} = .0167$$

Thus we would run LSD tests with an $\alpha_{pc} = .0167$, instead of .05.

2.3.1.1 Doing Bonferroni in R

emmeans will calculate LSD and correct the pvalues

```
# Fitted.Model<-emmeans(ANOVA.Table, ~Condition) #we already created this
pairs(Fitted.Model, adjust='bon')
```

```
## contrast          estimate      SE df t.ratio p.value
## Distrupted - Helpful      -3 0.7302967 12  -4.108  0.0044
## Distrupted - No          -1 0.7302967 12  -1.369  0.5880
## Helpful - No              2 0.7302967 12   2.739  0.0539
##
## P value adjustment: bonferroni method for 3 tests
```

2.3.2 Why is *a priori* Power affected?

Remember, *a priori* power is set by α_{pc} , Effect size and sample size. Let's examine power for the ANOVA and power for the follow-up tests (when we are changing the alpha)

For power analysis for One-way ANOVA, we will need f , which is what Cohen developed (like d for t-tests)

$$f = \sqrt{\frac{K-1}{K} \frac{F}{n}}$$

ES	Value	Size
f	.1	Small effect
f	.25	Medium effect
f	.4	Large effect

$f_{observed} = 1.08$, Large effect for our study. But remember, in small N studies these values might be hyperinflated. For now, we will believe it and use it calculate our *post-hoc power* (which is not predictive of our *a priori* power in the wild).

Note: You can approximate f from η^2 , $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$, but it will not be bias corrected as the formula above

```
library(pwr)
pwr.anova.test(k = 3, n = 5, f = 1.08, sig.level = 0.05, power = NULL)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##           k = 3
##           n = 5
##           f = 1.08
##      sig.level = 0.05
##           power = 0.9169511
##
## NOTE: n is number in each group
```

We can convert our f into d to try to estimate what overall effect size would be: $d = 2f$ (**Note:** this is only a conceptual formula; Cohen has different versions for different patterns of data)

$d = 2.16$

So at an $\alpha_{pc} .05$

```
pwr.t.test(n = 5, d = 2.16, power = NULL, sig.level = .05,
           type = c("two.sample"), alternative = c("two.sided"))
```

```
##
##      Two-sample t test power calculation
##
##           n = 5
##           d = 2.16
##      sig.level = 0.05
##           power = 0.8476039
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

We are above .80.

If we Bonferroni correct it $\alpha_{pc} .05/3 = .0167$

```
pwr.t.test(n = 5, d = 2.16, power = NULL, sig.level = .0167,
           type = c("two.sample"), alternative = c("two.sided"))
```

```
##
##      Two-sample t test power calculation
##
##           n = 5
##           d = 2.16
##      sig.level = 0.0167
##           power = 0.6569859
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

The power dropped! This could mean your ANOVA might find the effect, but your Bonferroni corrected tests might miss the effect. This is a very common thing that happens to people. I am often asked this question and the next questions they ask, but which is **real** the ANOVA or the Bonferroni tests?

2.3.3 Sidak method

Less conservative Bonferroni, but still very conservative.

$$\alpha_{pc} = 1 - (1 - \alpha_{EW})^{\frac{1}{j}}$$

Thus we would run tests with an $\alpha_{pc} = 1 - (1 - .05)^{\frac{1}{3}} = 0.017$, instead of .0167 (Bonferroni) or .05 (original value)

2.3.3.1 Doing Sidak in R

emmeans will calculate LSD and correct the pvalues

```
# Fitted.Model<-emmeans(ANOVA.Table, ~Condition) #we already created this
pairs(Fitted.Model, adjust='sidak')
```

```
## contrast          estimate      SE df t.ratio p.value
## Distrupted - Helpful      -3 0.7302967 12  -4.108  0.0043
## Distrupted - No          -1 0.7302967 12  -1.369  0.4803
## Helpful - No              2 0.7302967 12   2.739  0.0530
##
## P value adjustment: sidak method for 3 tests
```

2.3.4 Notes

There are many other types of correction, and these corrections are most often applied to pairwise approaches using the familywise philosophy.

3 Method of Contrast

Contrast as a way of customizing what you want to compare. It is the best way to do Confirmatory testing. You will not be able to apply HSD logic to it (as it's not pairwise approach, but you can do pairwise testing this way if you wanted too). This is a huge topic and we will cover only the basic types today and return to advanced forms later in the semester (such as orthogonal and polynomial testing).

3.1 Linear Contrast

This is basically an F-test, where you compare only the groups you want to compare on the fly [also allows you to merge groups on the fly].

$$F_{contrast} = \frac{MS_{contrast}}{MS_W}$$

Note: your book says:

$$F_{contrast} = \frac{SS_{contrast}}{MS_W}$$

That is correct because $df_{contrast}$ will always equal 1, because we can only compare 2 groups at once ($2 - 1 = 1$). Thus $MS_{contrast} = \frac{SS_{contrast}}{1}$. Mathematically, we can skip the $MS_{contrast}$ step because $MS_{contrast} = SS_{contrast}$

3.1.1 Step 1

Decide on how you want to weight (we call it C) your cells (they must add up to zero). I want to compare No + Disrupted VS Helpful

Cell	No Joint Action	Helpful Joint Action	Disrupted Joint Action
M	3	5	2
C	-1	2	-1

3.1.2 Step 2

Linear Sum (L)

$$L = \sum_{i=1}^k C_i M_i$$
$$L = -1 * 3 + 2 * 5 + -1 * 2 = 5$$

3.1.3 Step 3

Calculate the $MS/SS_{contrast}$

$$MS_{contrast} = \frac{SS_{contrast}}{1} = \frac{nL^2}{\sum C^2}$$
$$MS_{contrast} = \frac{SS_{contrast}}{1} = \frac{5 * 5^2}{(-1)^2 + (2)^2 + (-1)^2} = \frac{125}{6} = 20.833$$

3.1.4 Step 4

Calculate $F_{contrast}$

$$F_{contrast} = \frac{MS_{contrast}}{MS_W} = \frac{20.833}{1.33} = 15.625$$

Note: You can convert it to t-test like this

$$\sqrt{F_{contrast}} = t_{contrast}$$

$$\sqrt{15.625} = 3.953$$

3.1.5 Step 5

Look up F_{crit} using this df formula = $(1, df_w)$ or t_{crit} using this df formula = (df_w)

3.2 Doing Constrasts in R

- Check my the order of my fitted values

```
# Fitted.Model<-emmeans(ANOVA.Table, ~Condition) #we already created this
Fitted.Model
```

```
## Condition emmean      SE df lower.CL upper.CL
## Distrupted      2 0.5163978 12 0.8748659 3.125134
## Helpful          5 0.5163978 12 3.8748659 6.125134
## No               3 0.5163978 12 1.8748659 4.125134
##
## Confidence level used: 0.95
```

- Set up the hypothesis (make sure the order matches your fitted emmeans object!)

```
set1 <- list(
  H1 = c(-1,2,-1))
```

- Check: Do they each sum to zero?

```
sum(set1$H1)
```

```
## [1] 0
```

- Run the contrast

```
contrast(Fitted.Model, set1, adjust = "none")
```

```
## contrast estimate      SE df t.ratio p.value
## H1                5 1.264911 12   3.953  0.0019
```

It matches the by-hand calculation!

3.3 Doing Multiple Contrasts in R

- You can do a bunch in a family

```
set2 <- list(
  H1 = c(-1,2,-1),
  H2 = c(0,-1,1))
```

```
contrast(Fitted.Model, set2, adjust = "none")
```

```
## contrast estimate      SE df t.ratio p.value
## H1                5 1.264911 12   3.953  0.0019
## H2               -2 0.7302967 12  -2.739  0.0180
```

- You can correct this if you are worried about type I error (you cannot do Tukey here it will change it to Sidak. If you do 'none' = LSD result)

```
contrast(Fitted.Model, set2, adjust = "sidak")
```

```
## contrast estimate      SE df t.ratio p.value
## H1                5 1.264911 12   3.953  0.0038
## H2               -2 0.7302967 12  -2.739  0.0356
##
## P value adjustment: sidak method for 2 tests
```

4 Suggestions on Review Process

Follow-up testing is the hardest part of any analysis. There are often no correct answers and its the place reviewers will fight you the hardest.

Reviewers often have their own perspectives on what they want to see and how they want to see it. It could be because a) you misunderstood your own hypothesis, b) they misunderstood your hypothesis (maybe cause you explained it badly or they simply did not read), c) they demand to see all comparisons or they demand you remove all comparisons (reviewer 1 vs. reviewer 3 on the same paper!).

They may be ok with what you tested, but not how you tested it. a) They want Bonferroni correction only (because they want to control for type I error, want to kill your effect cause they don't like it [cynical I know], or they were taught LSD/HSD is the devil), b) they want specialized contrast approaches or pairwise approaches and you did the opposite, and c) they want you (or don't want you) to use the error term from the ANOVA (LSD/contrast approach; if you violated assumptions they might be right in specific cases but they need to explain their rational].

First, reviewers are not always right. However, you should assume YOU did not explain clearly what you did. The solution is to re-write carefully in plain English what you are testing, explain if it is "planned" or "unplanned" testing. Make sure to write simple sentences (technical writing, not English Lit. Make it clear **visually** as well as **verbally** what you are doing and why (guide the reader with graphs or tables). Cite sources that support your approach to follow-up testing (don't assume they know what you are doing and why). Explain in the review how you changed the wording and that you think the reviewer for helping to make it clearer.

Second, don't play games with the readers. If you $p = .07$ and it supports your hypothesis you call it significant in one place and later say $p = .06$ on another test is not-significant because it goes against your hypothesis. This is a huge red flag that you might be biased at best or a P-hacker at worst. You might be rejected without a revise for this, and the editor (and reviewers) might remember you in a bad way. (Most psych journals are not double blind).

Third, don't P-Hack. Don't hunt for a correction that makes your result significant when using a more conservative test fails. If you get that reviewer who is paying attention, they will eat you alive! It is that lack of trust that makes reviewers say Bonferroni everything. If you follow Confirmatory approaches, you can put the Exploratory stuff in the appendix and Bonferroni that. When a reviewer senses you are aware of P-hacking issues in that you explained how you are careful and not drawing too many conclusions they will trust you more.

Finally, reviewers might be right. Thank the reviewer for their careful read and correct your paper or retract it from the review process. Yes, it sucks but that's why we have reviewers. If they do their job well, everyone benefits!

5 APA Style Report

[Note: I cannot center and stuff in R, but you can knit to word and correct the formatting there]

When people perform joint actions, there is an increase the degree to which they trust each other (Launay, Dean & Bailes, 2013). Further, feelings of affiliation are increased when joint action occurs synchronously (Hove & Risen, 2009). In addition, we know that when know that when on member is not good at synchronizing it decreases feelings of social affiliation (Demos et al, 2017). Thus, we predict that synchronous joint action will increase feelings of trust over no joint action or that does not involve synchronous movements. However, what remains unclear is if the feelings of trust are the result of moving in time together (Hove & Risen, 2009) or simply the results of being in shared task (Demos et al, 2012). To test this, we designed a study three groups of participants were either in helpful joint action condition where a confederate tried to remain synchronous with their partner or the confederate was trying to do the joint action but in a disruptive (non-synchronous) way or was not engaged in joint action.

Methods

Procedure. Two people entered the lab (one a confederate) and the other a random participant. The status of the confederated was unknown the participant. They were asked to are asked to either a) move chairs independently, b)

Move a couch together, and confederate was good at following the leader c) Move a couch together and confederate hinders. Moving a couch requires the two people to coordinate their action. After the study ends, people rate how trustworthy their partner seems (1-7, with 7 = high trust).

Data Analysis. Analyses were conducted in R (3.5.1) with the afex package (Singmann et al., 2018) used to calculate the ANOVAs. The emmeans package (Lenth, 2018) was used for follow-up testing. Planned comparisons were followed up via linear contrasts. Unplanned contrasts were corrected were sidak corrected.

Results

A one-way between subjects ANOVA was conducted to compare the effect of joint action on feelings of trust in no, helpful, and disrupted joint action. There was a significant effect of joint action on feelings of trust, $F(2, 12) = 8.75$, $MS_W = 1.33$, $p < 0.01$, $\eta_g^2 = 0.59$. The planned contrast showed that the helpful joint actions was higher ($M = 5.00$, $SD = 1.22$) than combined no joint action and disrupted joint action ($M = 2.50$, $SD = 1.18$), $t(12) = 3.95$, $p = .0019$, $d = -2.06$. However, as seen in Figure 1 the planned contrast between the disrupted and no joint action did not yield a significant effect, $t(12) = 1.369$, $p = .20$, $d = 0.89$, but Sidak corrected unplanned contrasts yield significantly higher trust scores for helpful as compared to disrupted, $t(12) = 4.11$, $p = .003$, $d = 2.68$, and no joint action, $t(12) = 2.74$, $p = .04$, $d = -1.63$. This suggests that trust may be a results of the synchronous action as opposed to joint action.

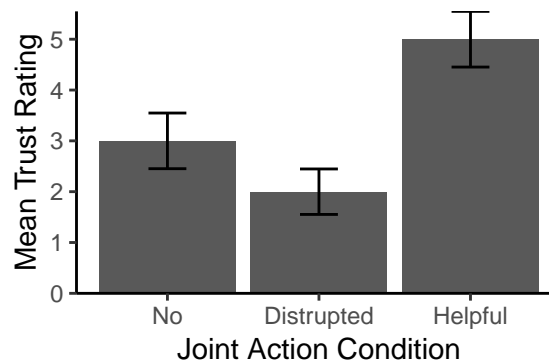


Figure 1. Mean levels of trust (1-7) by the three joint action conditions. Error bars are +/- 1 SE.

6 References

- Demos, A.P., Carter, D.J., Wanderley, M.M., & Palmer, C (2017). The unresponsive partner: Roles of social status, auditory feedback, and animacy in coordination of joint music performance. *Frontiers in Psychology*, 149(8).
- Demos, A.P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, 141(1), 49-53.
- Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social cognition*, 27(6), 949-960.
- Launay, J., Dean, R. T., & Bailes, F. (2013). Synchronization can influence trust following virtual interaction. *Experimental psychology*, 60(1), 53-63.
- Lenth, R. (2018). emmean: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.2.3. Available at: <https://CRAN.R-project.org/package=emmeans>.
- Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2018). afex: Analysis of Factorial Experiments. R Package Version 0.21-2. Available at: <https://CRAN.R-project.org/package=afex>.