Study_2

Farid Anvari and Daniel Lakens

Anonimize the data

```
# Anonimize data and store the analysis data file----
# # The code below is run on the raw non-anonymous data to generate the anonymous data file
# PANASdata <- read.csv(file('SESOI - btw person PA.csv'), stringsAsFactors = FALSE)
#
# #Save one case by fixing a clear participant typo (double 4 instead of single 4)
# PANASdata[69,15] <- "ID removed for anonimity"
#
# #mean PA for each participant
# PANASdata$mPA <- rowMeans(PANASdata[ , grep('PA_', colnames(PANASdata))], na.rm=TRUE)
#
# #create new variable for each participant that extracts mean PA from their partner
# # based on StudentID (from other observations in the dataset) that matches the partnerID
# # of the target participant
# PANASdata$mPApartner <- PANASdata[match(PANASdata$PartnerID,
#                                       PANASdata$StudentID, nomatch = NA_integer_,
#                                       incomparables = NULL),
#                                 'mPA']
#
# # Get all PANAS Data from partner (not just the mean)
# PANAS_partner <- PANASdata[match(PANASdata$PartnerID,
#                           PANASdata$StudentID,
#                           nomatch = NA_integer_,
#                           incomparables = NULL),
#                     c(5:14)]
# # Change column names
# colnames(PANAS_partner)[1:10] <- paste(colnames(PANAS_partner)[1:10], "partner", sep = "_")
# # Combine the 2 dataframes
# PANASdata <- cbind(PANASdata,PANAS_partner)
#
# # Below is a check of we get the same means - and we do
# PANASdata$mPA2 <- rowMeans(PANASdata[ ,19:28], na.rm=TRUE)
#
# # We can no delete columns to make the data anonymous
# PANASdata <- subset(PANASdata, select = -c(StudentID, PartnerID, mPA2))
#
# # Add random StudentID as participant identifier
# PANASdata$StudentID <- c(1:nrow(PANASdata))
#
# #Then write the csv file
# write.csv2(PANASdata, file = "PANAS_study2_data_anon.csv", row.names = FALSE)
```

Read in raw data

```r
# Read in data----
PANASdata <- read.csv2(file('PANAS_study2_data_anon.csv'), stringsAsFactors = FALSE)

#remove any observations with NA values
PANASdata <- PANASdata[complete.cases(PANASdata), ]
```

Calculate Difference Scores

This uses a custom function written to calculate effect sizes and confidence interval for within designs. The code isfor Study 2 greatly overlaps with the code used in Study 1.

```r
#calculate the difference in mean PA between the pairs,
PANASdata$PA_diff <- PANASdata$mPA - PANASdata$mPApartner

#Calculate all statistics for each global anchor question group.----

sum_data<-data.frame(t = numeric(0),
                     df = numeric(0),
                     p = numeric(0),
                     ci_lower_mean = numeric(0),
                     ci_upper_mean = numeric(0),
                     mean_1 = numeric(0),
                     mean_2 = numeric(0),
                     m_diff = numeric(0),
                     ci_l_m_diff = numeric(0),
                     ci_u_m_diff = numeric(0),
                     d_av = numeric(0),
                     d_av_unb = numeric(0),
                     s_av = numeric(0),
                     s_diff = numeric(0),
                     ci_l_d_av = numeric(0),
                     ci_u_d_av = numeric(0),
                     d_z = numeric(0),
                     d_z_unb = numeric(0),
                     ci_l_d_z = numeric(0),
                     ci_u_d_z = numeric(0),
                     N = numeric(0),
                     m1 = numeric(0),
                     m2 = numeric(0),
                     sd1 = numeric(0),
                     sd2 = numeric(0),
                     cor = numeric(0),
                     vi = numeric(0))
for(i in 1:5){
  temp <- t.test(PANASdata$PA_diff[PANASdata$CompareQ == i])
  sum_data[i,1] <- temp$statistic
  sum_data[i,2] <- temp$parameter
  sum_data[i,3] <- temp$p.value
  sum_data[i,4] <- temp$conf.int[1]
  sum_data[i,5] <- temp$conf.int[2]
  sum_data[i,6] <- temp$estimate[1]
  sum_data[i,7] <- temp$estimate[2]
    temp <- effect_size_d_paired(PANASdata$mPApartner[PANASdata$CompareQ == i],
```

```
                                PANASdata$mPA[PANASdata$CompareQ == i])
  sum_data[i,8] <- temp$m_diff
  sum_data[i,9] <- temp$ci_l_m_diff
  sum_data[i,10] <- temp$ci_u_m_diff
  sum_data[i,11] <- temp$d_av
  sum_data[i,12] <- temp$d_av_unb
  sum_data[i,13] <- temp$s_av
  sum_data[i,14] <- temp$s_diff
  sum_data[i,15] <- temp$ci_l_d_av
  sum_data[i,16] <- temp$ci_u_d_av
  sum_data[i,17] <- temp$d_z
  sum_data[i,18] <- temp$d_z_unb
  sum_data[i,19] <- temp$ci_l_d_z
  sum_data[i,20] <- temp$ci_u_d_z
  sum_data[i,21] <- temp$N
  sum_data[i,22] <- temp$m1
  sum_data[i,23] <- temp$m2
  sum_data[i,24] <- temp$sd1
  sum_data[i,25] <- temp$sd2
  sum_data[i,26] <- temp$cor
  sum_data[i,27] <- 1/sum_data$N[i] + sum_data$d_z[i]^2 / (2*sum_data$N[i])
  #this is the formula for the variance based on the metafor package.
  #It leads to CI slightly different from dz_l and dz_u because
  #these are based on non-central t (using MBESS package).
}
```

```
## Mean Difference     : -0.4 95% CI [-4.0594;3.2594]
## Cohen's d_z_unb    : -0.1552 95% CI [-1.7153;1.1216]
## Cohen's d_av_unb   : -0.3019 95% CI [-2.7246;1.7817]
## s_diff: 1.4731, s_av: 0.7572, sd1: 0.5859, sd2: 0.8963, cor: -0.9743
## N = 3 pairs.Mean Difference    : -0.3952 95% CI [-0.7108;-0.0797]
## Cohen's d_z_unb    : -0.5485 95% CI [-1.0518;-0.1043]
## Cohen's d_av_unb   : -0.6669 95% CI [-1.2479;-0.1237]
## s_diff: 0.6932, s_av: 0.5702, sd1: 0.4711, sd2: 0.6544, cor: 0.2753
## N = 21 pairs.Mean Difference     : 0.0349 95% CI [-0.1478;0.2176]
## Cohen's d_z_unb    : 0.0577 95% CI [-0.2436;0.3618]
## Cohen's d_av_unb   : 0.0665 95% CI [-0.2774;0.412]
## s_diff: 0.5936, s_av: 0.5151, sd1: 0.5273, sd2: 0.5026, cor: 0.3365
## N = 43 pairs.Mean Difference     : 0.1781 95% CI [-0.085;0.4413]
## Cohen's d_z_unb    : 0.2381 95% CI [-0.1113;0.6034]
## Cohen's d_av_unb   : 0.3159 95% CI [-0.1454;0.788]
## s_diff: 0.7299, s_av: 0.5501, sd1: 0.6069, sd2: 0.4867, cor: 0.1227
## N = 32 pairs.Mean Difference     : 1.15 95% CI [-10.9209;13.2209]
## Cohen's d_z_unb    : 0 95% CI [-1.3438;3.5185]
## Cohen's d_av_unb   : 0 95% CI [-1.8378;4.8121]
## s_diff: 1.3435, s_av: 0.6946, sd1: 0.495, sd2: 0.8485, cor: -1
## N = 2 pairs.
```
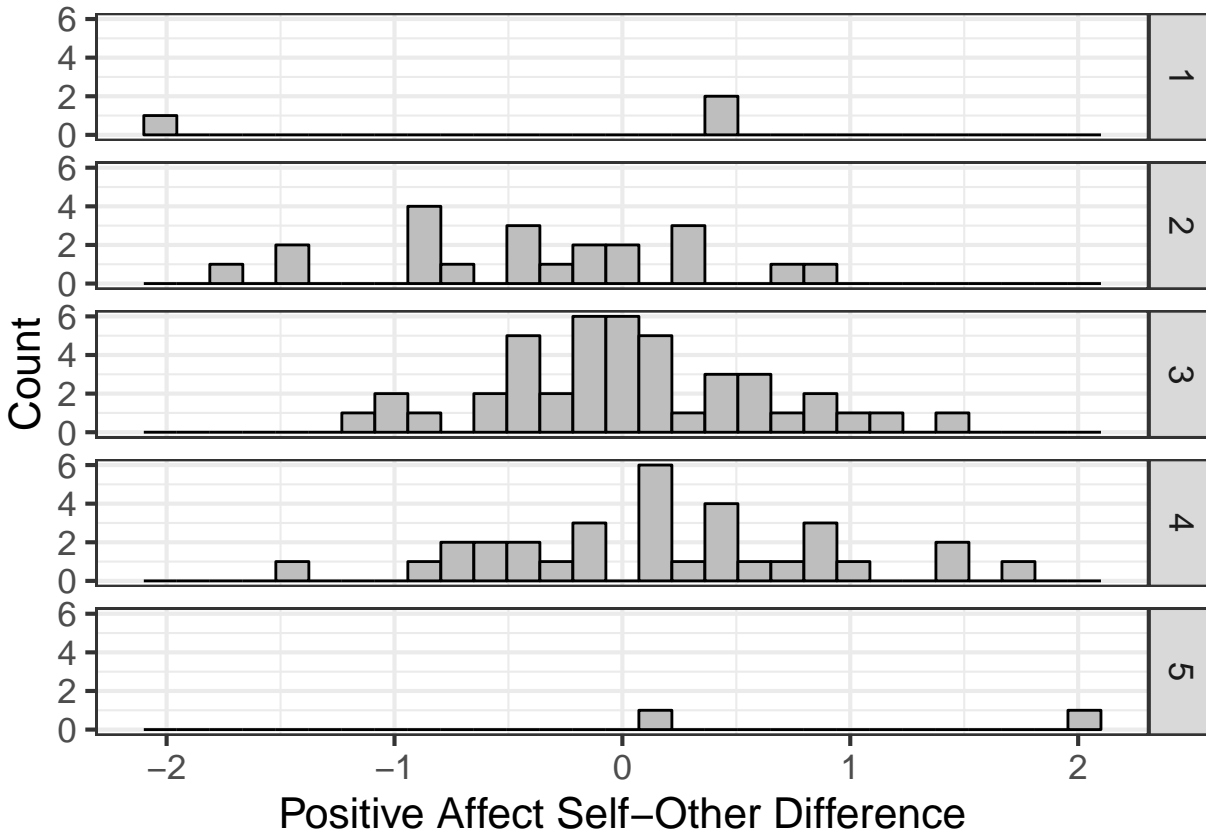
Plot the distributions

Note that these plots are not in the manuscript.

```
#Plot the data----
#For positive
ggplot(PANASdata, aes(PA_diff)) +
  geom_histogram(colour = "black", fill = "grey") +
  xlab("Positive Affect Self-Other Difference") + ylab("Count") +
  theme_bw(base_size = 16) +
  facet_grid(rows = vars(PANASdata$CompareQ))
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Summary Stats Table

The stats are only for the positive affect conditions (who answered 1 to 5 on the global transition question). We did not include questions about negative affect in this study.

```
#Create summary stats table
round(sum_data[,c(21,8,9,10,17,19,20,11,15,16)],2)
```

```
##     N m_diff ci_l_m_diff ci_u_m_diff   d_z ci_l_d_z ci_u_d_z  d_av
## 1   3  -0.40       -4.06        3.26 -0.27    -1.72     1.12 -0.53
## 2  21  -0.40       -0.71       -0.08 -0.57    -1.05    -0.10 -0.69
## 3  43   0.03       -0.15        0.22  0.06    -0.24     0.36  0.07
## 4  32   0.18       -0.09        0.44  0.24    -0.11     0.60  0.32
## 5   2   1.15      -10.92       13.22  0.86    -1.34     3.52  1.66
##   ci_l_d_av ci_u_d_av
## 1     -2.72      1.78
```

```
## 2      -1.25     -0.12
## 3      -0.28      0.41
## 4      -0.15      0.79
## 5      -1.84      4.81
```

We add a TOTAL line in the summary graph, which is based on combining the 2 little change groups (the people who answer 2 and 4). For this, we need to recode the data in one group (otherwise the differences in both groups which are in opposite directions cancel each other out). Then we calculate the means and sd, and perform a t-test in R to get the CI.

```
#And the combined means for the little change individuals:
PANASdata$PA_change_rec <- PANASdata$mPA - PANASdata$mPApartner

#Need to flip around the scores from group 2 (which are in the opposite direction)
PANASdata$PA_change_rec <- ifelse(PANASdata$CompareQ == 2 | PANASdata$CompareQ == 1,
                                  PANASdata$PA_change_rec * -1,
                                  PANASdata$PA_change_rec * 1)

#Mean only for little change group: positive
mean(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4])
```

```
## [1] 0.2641509
```

```
sd(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4])
```

```
## [1] 0.7168559
```

```
t.test(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4])
```

```
##
##  One Sample t-test
##
## data:  PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ ==    4]
## t = 2.6826, df = 52, p-value = 0.00977
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.06656098 0.46174091
## sample estimates:
## mean of x
## 0.2641509
```

Meta-Analytic Approach

As for study 1, although we originally planned (and pre-registered) to simply follow the 'combine the 2 slightly changed groups' analysis approach, upon consideration and after looking at the data from both studies, we changed our mind. Combining these 2 groups assumes 1) the effect sizes are similar, and 2) the standard deviations are similar. That might be true, but we believe for now it is too strong an assumption to make. An alternative solution is to not simply combine the two groups, but to meta-analyze them using a random effects meta-analysis. We can explore heterogeneity, and after several datasets have been collected, decide if effect sizes and sd's are similar or not.

```
# We can perform a meta-analysis based on this data.
# We can use the escalc function - which reports a bias corrected version of dz
#(so the estimates are slightly lower)
g <- escalc(ni = sum_data$N,
            m1i = sum_data$m1,
```

```
            m2i = sum_data$m2,
            sd1i = sum_data$sd1,
            sd2i = sum_data$sd2,
            ri = sum_data$cor,
            measure = "SMCC")
# We reverse the direction of the effect for individuals in group 2 by using abs(yi)
result <- rma(abs(yi), vi, data=g[c(2,4),])
result
```
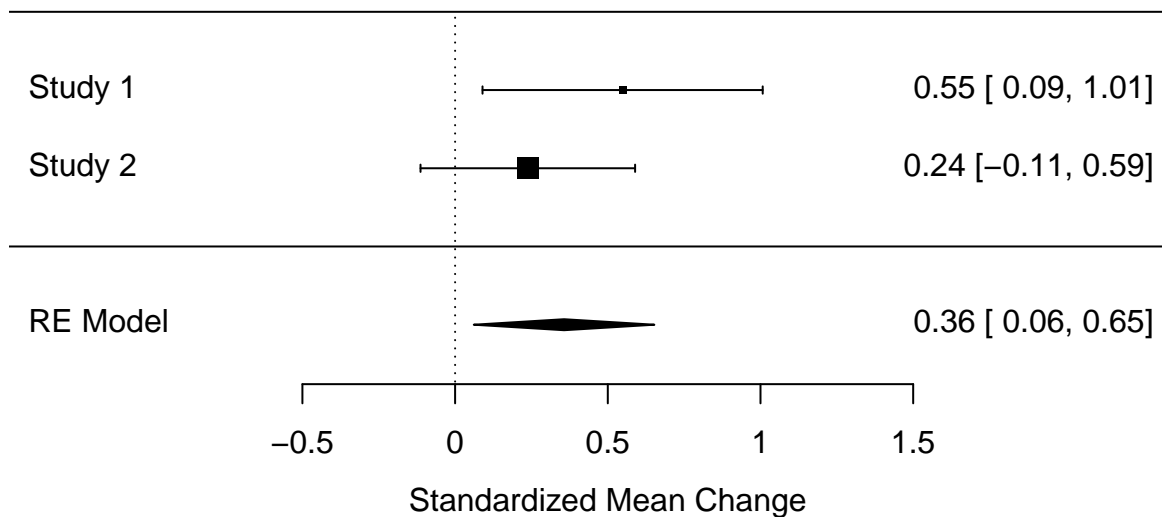
```
##
## Random-Effects Model (k = 2; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.0047 (SE = 0.0681)
## tau (square root of estimated tau^2 value):      0.0687
## I^2 (total heterogeneity / total variability):   9.79%
## H^2 (total variability / sampling variability):  1.11
##
## Test for Heterogeneity:
## Q(df = 1) = 1.1086, p-val = 0.2924
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.3568  0.1509  2.3652  0.0180  0.0611  0.6525  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
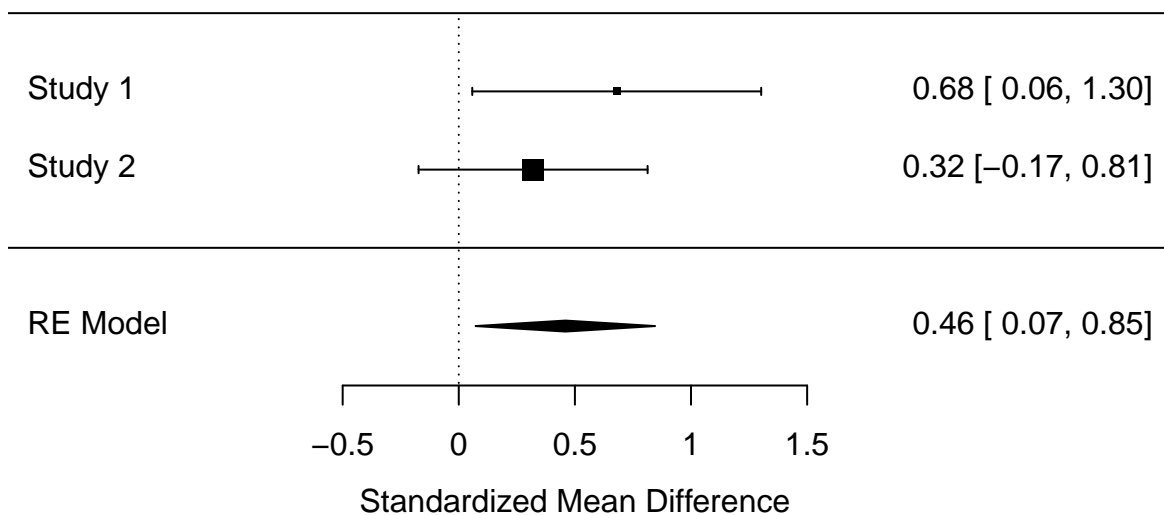```

```
forest(result)
```

```
#We can also meta-analyze d_av (treating the 2 scores as independent)----
# For positive
g <- escalc(n1i = sum_data$N,
            n2i = sum_data$N,
            m1i = sum_data$m1,
            m2i = sum_data$m2,
            sd1i = sum_data$sd1,
            sd2i = sum_data$sd2,
            measure = "SMD")
# We reverse the direction of the effect for individuals in group 2 by using abs(yi)
result <- rma(abs(yi), vi, data=g[c(2,4),])
result
```

```
##
## Random-Effects Model (k = 2; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.1160)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  1.00
##
## Test for Heterogeneity:
## Q(df = 1) = 0.7910, p-val = 0.3738
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
```

```
##   0.4589  0.1972  2.3274  0.0199  0.0724  0.8453  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
forest(result)
```



```r
#Note that the CI for d_z differ slightly from the meta-analytic effect sizes
#(due to the use of non-central t in MBESS and bias correction)
sum_data[c(2,4),c(18:20)]
```

```
##      d_z_unb   ci_l_d_z   ci_u_d_z
## 2 -0.5485413 -1.0517900 -0.1042967
## 4  0.2380930 -0.1113298  0.6033618
```

Note the heterogeneity is substantial, increasing our doubt that simply combining these two groups is the best analytic approach.

Plot

Then the plot is produced (when saving, the resolution is better)

```r
#Create a graph showing individual change

longPA = PANASdata
longPA$PAself = longPA$mPA #new variable with more descriptive name
longPA$PApartner = longPA$mPApartner #new variable with more descriptive name
```

```r
#WHEN WE MAKE IT LONG FORMAT and we only want those in the little changed
#PA group first
longPA = melt(longPA[longPA$CompareQ==2 | longPA$CompareQ == 4,],
              measure.vars = c("PAself", "PApartner") )

#looking at the dataframe created, can see that each participant now has a
# T1_PA, and T2_PA entry (i.e., in long format)

longPA$CompareQ = as.factor(longPA$CompareQ)
levels(longPA$CompareQ) <- c("a little less positive", "a little more positive")
#renaming the levels

levels(longPA$variable) <- c("PA Self", "PA Partner") #renaming the levels

plot1 <- ggplot(data = longPA, aes(x = variable, y = value)) +
  geom_point(size = 2, color = "black", alpha = 0.3) + #colour points by globalPA
  geom_path(aes(group = StudentID, alpha = 0.3)) + #spaghetti plot
  ylab("Mean Positive Affect") +
  xlab("Time") +
  theme_bw(base_size = 18) +
  facet_grid(~ CompareQ) +
  theme(legend.position="none") +
  scale_x_discrete(labels=c("T1_PA" = "Time 1", "T2_PA" = "Time 2")) +
  theme(axis.title.x=element_blank(),
        strip.background =element_rect(fill="white"),
        panel.grid.major.x = element_blank())
plot1
```
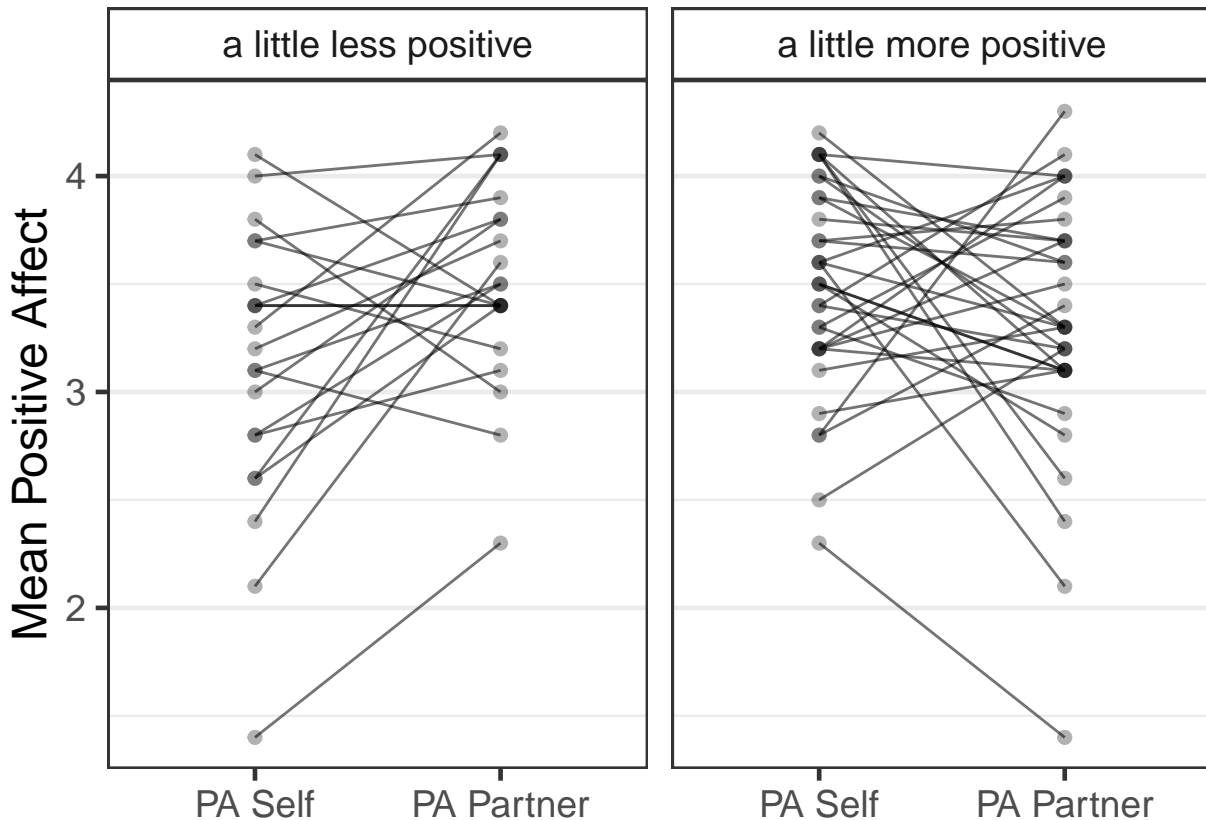
```
#code to save the plot
jpeg(file="plot1_study2.jpg",width=2000,height=1400, units = "px", res = 300)
plot1
dev.off()
```

```
## pdf
##   2
```

This is the end of the results reported in the main text.

Exploratory analyses not reported in the manuscript.

```
### test of difference between the "little-changed" groups and the "same" group
#for PA
t.test(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4],
       PANASdata$PA_change_rec[PANASdata$CompareQ == 3])
```

```
##
##  Welch Two Sample t-test
##
## data:  PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ ==  and PANASdata$PA_chan
## t = 1.7141, df = 93.952, p-value = 0.0898
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03630164  0.49483609
## sample estimates:
```

```
##  mean of x  mean of y
## 0.26415094 0.03488372
```

```r
mean(PANASdata$PA_change_rec[PANASdata$CompareQ == 3])
```

```
## [1] 0.03488372
```

```r
sd(PANASdata$PA_change_rec[PANASdata$CompareQ == 3])
```

```
## [1] 0.5935702
```

```r
cohen.d(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4],
        PANASdata$PA_change_rec[PANASdata$CompareQ == 3],
        pooled=TRUE,paired=FALSE,
        na.rm=TRUE, hedges.correction=FALSE,
        conf.level=0.95)
```

```
##
## Cohen's d
##
## d estimate: 0.3449686 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.06552873  0.75546586
```

Exploratory analyses not reported in the manuscript

Analyzing Individual Results.

The code below calculates the numbers not in the main text:

"As in Study 1, we conducted exploratory analyses examining individual variability. In total, 34 of the 53 participants who reported feeling a little less or a little more positive than their partner had a difference score on the PA dimension of the PANAS in the direction they estimated, 2 showed no difference from their partner, and 17 had difference scores in the opposite direction to that which they indicated on the comparison question (e.g., they said they felt a little more positive than their partner when their mean PA was actually lower than their partner's)."

```r
#looking at individual variability
sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4] >= 0.26)
```

```
## [1] 26
```

```r
# 26/53 in the little-different group showed differences at MID level or higher

sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4] > 0)
```

```
## [1] 34
```

```r
# 34/53 showed differences greater than zero scale points
# in the direction they said they changed on the global Qs

sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4] == 0)
```

```
## [1] 2
```

```r
# 2/53 showed differences equal to zero scale points
# in the direction they said they changed on the global Qs
```

```r
sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 2 | PANASdata$CompareQ == 4] < 0)
```

```
## [1] 17
```

```r
# 17/53 showed differences in the opposite direction to what they said
# (e.g., said they felt a little more positive emotions but PA scores were lower)

sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 2] < 0)
```

```
## [1] 5
```

```r
#5/21 of those who felt less positive showed change in opposite direction

sum(PANASdata$PA_change_rec[PANASdata$CompareQ == 4] < 0)
```

```
## [1] 12
```

```r
#12/32 of those who felt more positive showed change in opposite direction

#####
#testing difference between those who were a little less positive and those who
# were a little more positive
t.test(PANASdata$PA_change_rec[PANASdata$CompareQ == 2],
       PANASdata$PA_change_rec[PANASdata$CompareQ == 4])#difference is nonsig
```

```
##
##  Welch Two Sample t-test
##
## data:  PANASdata$PA_change_rec[PANASdata$CompareQ == 2] and PANASdata$PA_change_rec[PANASdata$Compare
## t = 1.092, df = 44.495, p-value = 0.2807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1834477  0.6176739
## sample estimates:
## mean of x mean of y
## 0.3952381 0.1781250
```

```r
#difference is nonsig

#a little less positive vs the same
t.test(PANASdata$PA_change_rec[PANASdata$CompareQ == 2],
       PANASdata$PA_change_rec[PANASdata$CompareQ == 3])
```

```
##
##  Welch Two Sample t-test
##
## data:  PANASdata$PA_change_rec[PANASdata$CompareQ == 2] and PANASdata$PA_change_rec[PANASdata$Compare
## t = 2.0443, df = 34.766, p-value = 0.04856
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.002407885 0.718300863
## sample estimates:
##  mean of x  mean of y
## 0.39523810 0.03488372
```

```r
#a little more positive vs the same
t.test(PANASdata$PA_change_rec[PANASdata$CompareQ == 4],
```

```
        PANASdata$PA_change_rec[PANASdata$CompareQ == 3])
```

```
## 
##  Welch Two Sample t-test
## 
## data:  PANASdata$PA_change_rec[PANASdata$CompareQ == 4] and PANASdata$PA_change_rec[PANASdata$Compare
## t = 0.90882, df = 58.555, p-value = 0.3672
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1721894  0.4586720
## sample estimates:
##  mean of x  mean of y
## 0.17812500 0.03488372
```