Supplement: CogSci 2016 Paper

written by: Florian Sense

January 2016

Contents

Abstract

A PDF of the submitted paper is included in this repository.

Abstract:

> Adaptive fact learning systems have been developed to make optimal use of testing and spacing effects by taking into account individual differences in learning efficiency. Measures derived from these systems, capturing the individual differences, predict later performance in similar and different fact learning tasks. Additionally, there is a rich body of literature showing that individual differences in general cognitive ability or working memory capacity are predictive achievement tests. If these measures also influence fact learning, incorporating them might further enhance adaptive systems. However, here we show that performance during fact learning is neither related to working memory capacity nor general cognitive ability. This means that the individual differences captured by our adaptive learning system encapsulate characteristics of learners that are independent of their general cognitive ability. Consequently, adaptive learning methods should focus primarily on memory-related processes.

The Data

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```r
source("functions/print.correlation.R")
source("functions/clean_graphs.R")
data <- read.csv("data/data for cogsci paper.csv")
head(data)
```

```
##   X.1  X subject           timestamp male     alpha   condition
## 1   4  4   68086 10/12/2015 17:02:50    0 0.3271115 slimstampen
## 2   5  5   68752 10/12/2015 17:02:52    0 0.2995743 slimstampen
## 3   6  6   68614 10/12/2015 17:02:55    0 0.2933250 slimstampen
## 4   8  8   68464 10/12/2015 17:03:02    1 0.3400560 slimstampen
## 5  10 10   68476 10/12/2015 17:03:15    0 0.2044200 slimstampen
## 6  11 11   68392 10/12/2015 17:03:22    0 0.2459314 slimstampen
##   ospan.total rotspan.total symspan.total test1 test2         iq
## 1          56            29            NA    21    18 -0.1643746
## 2          59            31            25    32    26 -1.8932232
## 3          43            29            27    24    14 -1.7751339
## 4          64            27            37    18    12 -0.1035977
## 5          60            31            20    35    35 -0.7585407
## 6          57            36            32    35    35 -0.5359615
```

```r
length(unique(data$subject)) # number of participants
```

```
## [1] 42
```

Demographic Information

A summary of the demographic information reported in the paper.

```r
demo <- read.csv("data/demographics.csv")
colnames(demo) <- c("row", "timestamp", "subj", "yob", "gender", "native", "Swahili")
levels(demo$gender) <- c("male", "female")
demo$age <- 2015 - demo$yob
```

Information about age

```r
median(demo$age)
```

```
## [1] 19
```

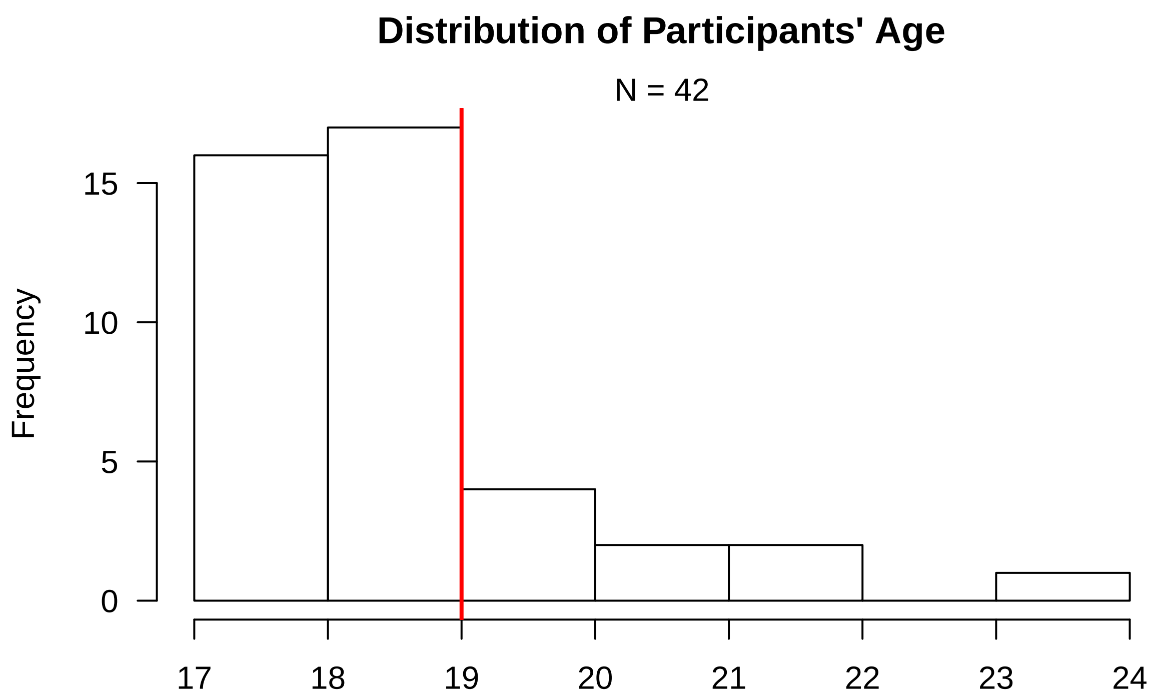
```r
sd(demo$age)
```

```
## [1] 1.342593
```

```r
range(demo$age)
```

```
## [1] 17 24
```

```r
hist(demo$age, main="Distribution of Participants' Age", xlab="", las=1)
mtext(paste("N =", nrow(demo)))
abline(v=median(demo$age), lwd=2, col="red") # indicate the median age on the plot
```



Information about gender

```
table(demo$gender)
```

```
##
##   male female
##     28     14
```

```
mean(demo$gender == "female") * 100 # percent female
```

```
## [1] 33.33333
```

Data reported in Table 1

Table 1 reports means, standard deviations, and the range for the three complex span tasks and their composite score. Also reported are the correlations.

```
# Compute the composite score:
data$wmc <- rowMeans(cbind(scale(data$ospan.total),
                           scale(data$rotspan.total),
                           scale(data$symspan.total)), na.rm = TRUE)

tmp <- subset(data,
              select = c("ospan.total", "rotspan.total", "symspan.total", "wmc"))
```

Compute the summary statistics in Table 1:

```r
apply(tmp, 2, mean, na.rm=TRUE)
```

```
##   ospan.total rotspan.total symspan.total           wmc
##  59.500000000  28.952380952  31.439024390  -0.001478521
```

```r
apply(tmp, 2, sd, na.rm=TRUE)
```

```
##   ospan.total rotspan.total symspan.total           wmc
##      9.232102      7.298354      6.942798      0.844579
```

```r
apply(tmp, 2, range, na.rm=TRUE)
```

```
##      ospan.total rotspan.total symspan.total       wmc
## [1,]          39            10            14 -1.642233
## [2,]          75            40            42  1.246696
```

Correlations among the three complex span tasks:

```r
print.correlation(tmp$ospan.total, tmp$rotspan.total, "OSpan", "RotSpan")
```

```
## cor(OSpan, RotSpan) = 0.56 with t(40) = 4.27 and p >= 1e-04.
##     The BF is 189.75 (in favor of alt.).
```

```r
print.correlation(tmp$ospan.total, tmp$symspan.total, "OSpan", "SymSpan")
```

```
## cor(OSpan, SymSpan) = 0.635 with t(39) = 5.13 and p >= 0.
##     The BF is 2327.53 (in favor of alt.).
```

```r
print.correlation(tmp$rotspan.total, tmp$symspan.total, "RotSpan", "SymSpan")
```

```
## cor(RotSpan, SymSpan) = 0.541 with t(39) = 4.02 and p >= 3e-04.
##     The BF is 90.49 (in favor of alt.).
```

Correlations with the composite score:

```r
print.correlation(tmp$wmc, tmp$ospan.total, "WMC", "OSpan")
```

```
## cor(WMC, OSpan) = 0.863 with t(40) = 10.82 and p >= 0.
##     The BF is 50063052460.06 (in favor of alt.).
```

```r
print.correlation(tmp$wmc, tmp$rotspan.total, "WMC", "RotSpan")
```

```
## cor(WMC, RotSpan) = 0.826 with t(40) = 9.28 and p >= 0.
##     The BF is 691068558.28 (in favor of alt.).
```

```r
print.correlation(tmp$wmc, tmp$symspan.total, "WMC", "SymSpan")
```

```
## cor(WMC, SymSpan) = 0.854 with t(39) = 10.25 and p >= 0.
##     The BF is 8005776935.47 (in favor of alt.).
```

All *p*-values for the correlations are below .05 and the corresponding Bayes factors are reported as well. The Bayes factors for the composite score (WMC) are ridiculously large. So large, in fact, that the numbers are not very meaningful which is why I reported them in the paper as "well over one billion". (They are, in fact, over 9 quadrillion!)

We also report the correlation between the two test scores in the paper:

```r
print.correlation(data$test1, data$test2, "Test 1", "Test 2")
```

```
## cor(Test 1, Test 2) = 0.881 with t(39) = 11.64 and p >= 0.
##     The BF is 305419269898.42 (in favor of alt.).
```

Variation in estimated parameters

The *rate of forgetting* is the mean parameter value across all items for each participant.

```r
param <- read.csv("data/parameters.csv")

# Compute the rate of forgetting:
rof <- aggregate(alpha ~ subj, param, mean)

mean(rof$alpha)
```

```
## [1] 0.2880531
```

```r
sd(rof$alpha)
```

```
## [1] 0.04580058
```

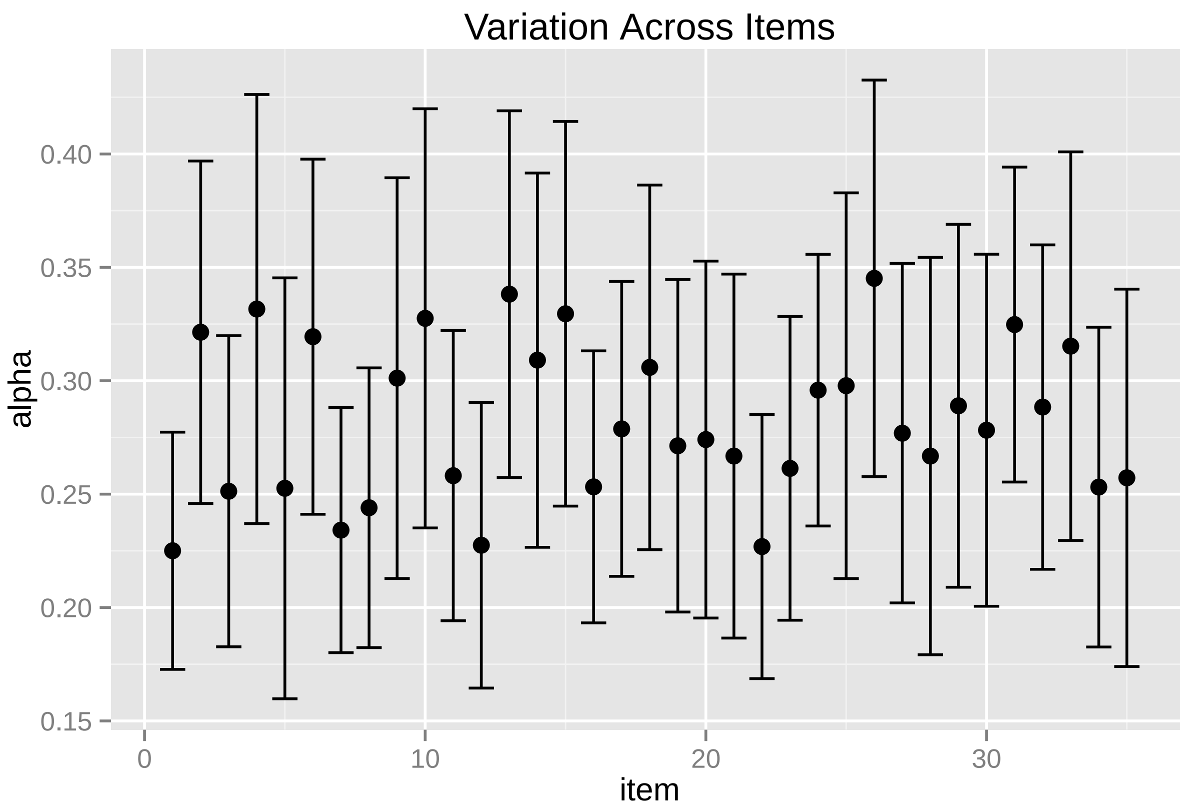
```r
range(rof$alpha)
```

```
## [1] 0.1856343 0.4044937
```

```r
perItem <- aggregate(alpha ~ item, param, mean)
perItem <- cbind( perItem, sd = aggregate(alpha ~ item, param, sd)[, 2] )

perSubj <- aggregate(alpha ~ subj, param, mean)
perSubj <- cbind( perSubj, sd = aggregate(alpha ~ subj, param, sd)[, 2] )

ggplot(perItem, aes(x=item, y=alpha)) + geom_errorbar(aes(ymin=alpha-sd, ymax=alpha+sd)) +
  geom_point(size=3) + ggtitle("Variation Across Items")
```




```
ggplot(perSubj, aes(x=subj, y=alpha)) + geom_errorbar(aes(ymin=alpha-sd, ymax=alpha+sd)) +
  geom_point(size=3) + ggtitle("Variation Across Participants")
```

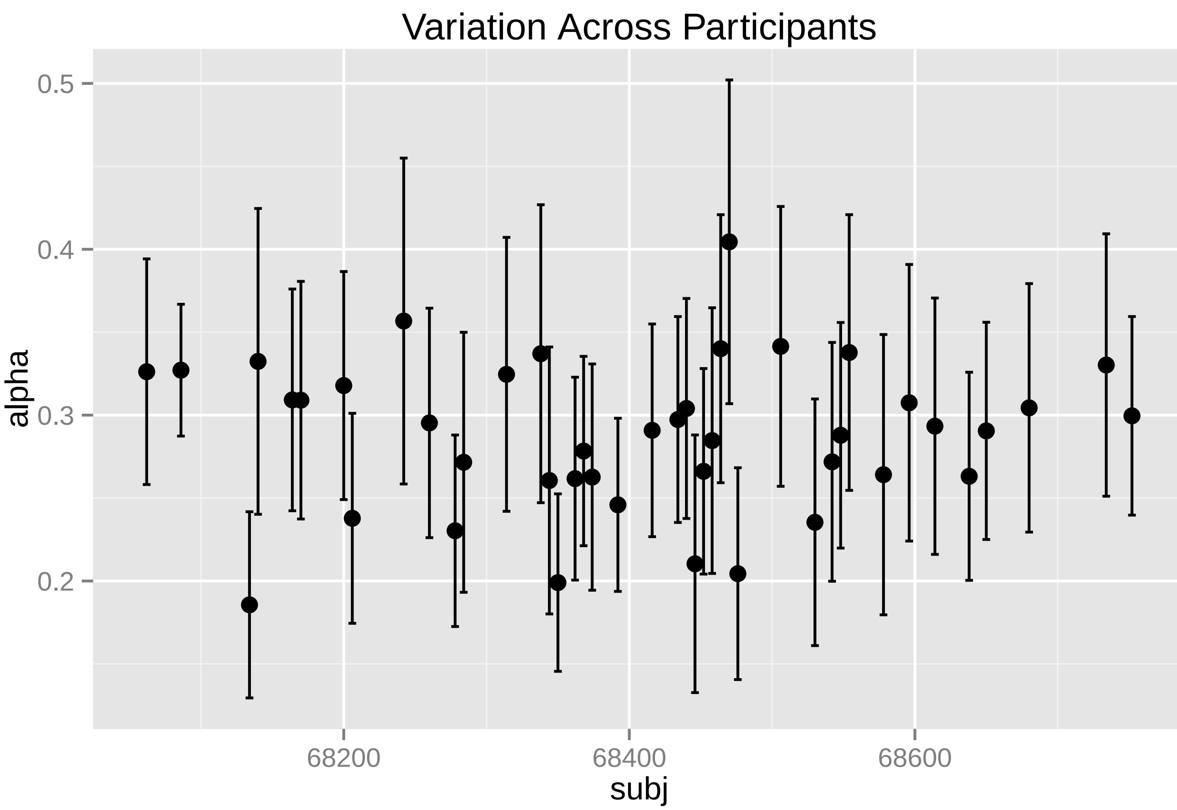


Figure 1

The figure is a bit bigger so I'll save it to a seperate file.

```
pdf("Figure 1.pdf", 7, 7)
subset <- cbind(data$alpha, data$test2, data$iq, data$wmc)
colors <- rep("#90AFC5", ncol(subset))
variables <- c("Rate of Forgetting", "Second Test", "GCA", "WMC")
clean.overview(subset, variables, colors)
dev.off()
```

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

As stated in the paper, these results are not very different if the test scores from the first test (instead of those from the second test) are used.

```
pdf("Figure 1 with scores from first test.pdf", 7, 7)
subset <- cbind(data$alpha, data$test1, data$iq, data$wmc)
variables <- c("Rate of Forgetting", "First Test", "GCA", "WMC")
clean.overview(subset, variables, colors)
dev.off()
```

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

Session Information

Information regarding the version of R and all loaded packages when compiling this document.

```
print(sessionInfo(), locale = FALSE)
```

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-apple-darwin10.8.0 (64-bit)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] plotrix_3.5-11 ggplot2_1.0.1
##
## loaded via a namespace (and not attached):
##  [1] MASS_7.3-35      Rcpp_0.11.6      colorspace_1.2-6 digest_0.6.8
##  [5] evaluate_0.8     grid_3.1.2       gtable_0.1.2     htmltools_0.2.6
##  [9] knitr_1.12       labeling_0.3     munsell_0.4.2    plyr_1.8.2
## [13] proto_0.3-10     reshape2_1.4.1   rmarkdown_0.9.2  scales_0.2.4
## [17] stringr_0.6.2    tools_3.1.2      yaml_2.1.13
```