Exploratory analyses of Wimber et al 2015

Kevin Potter

May 13, 2016

This is a report on an assortment of generalized linear models I've explored so far for fitting the data from the final recognition memory test of Wimber et al. (2015). Much of this is about determining the appropriate predictors to use in our pre-registration attempt. For speed and convenience, I've currently restricted myself to a frequentist framework.

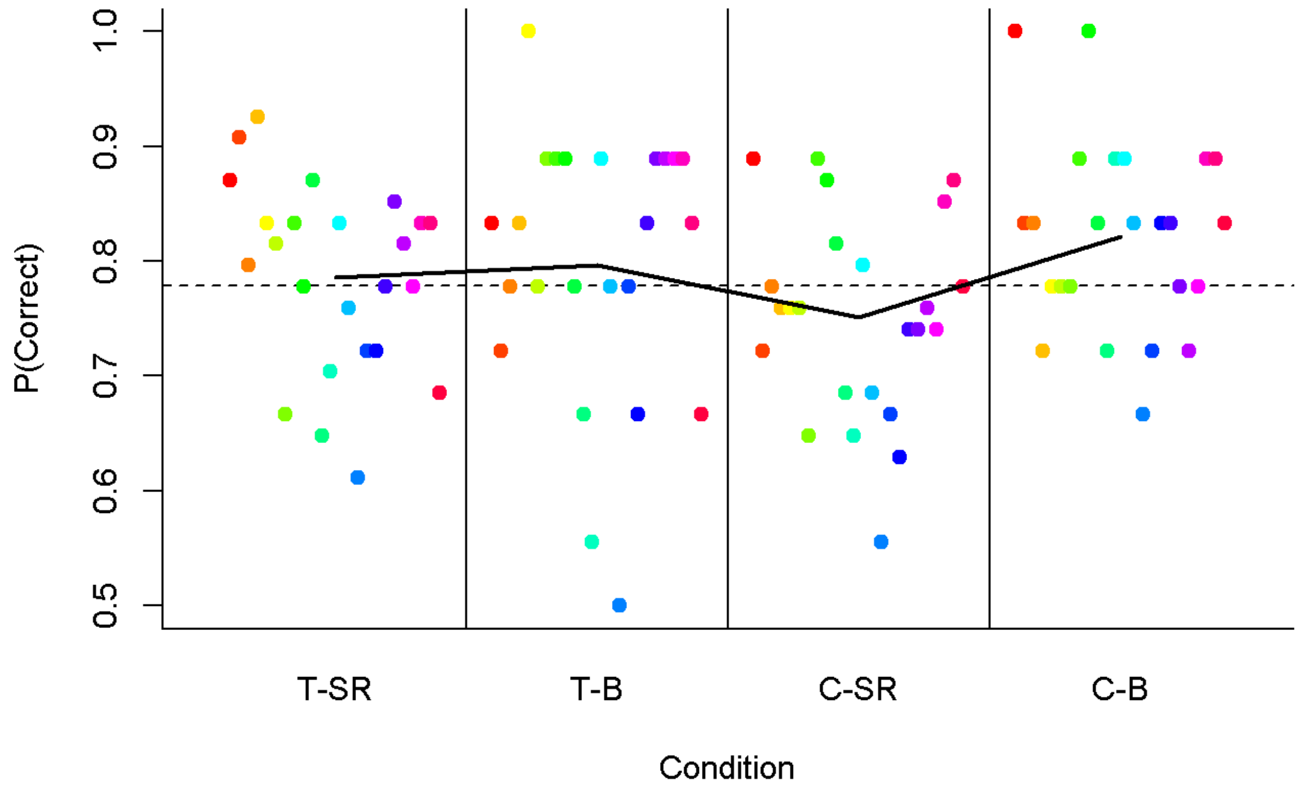
```
# Load in package for fitting mixed generalized linear models
library( lme4 )
```

```
## Loading required package: Matrix
```

The initial predictors I worked with were limited to the 4 conditions in the final test and subject/item factors.

```
d = as.data.frame( AD$Accuracy ) # Dependent variable
colnames(d) = 'Y'
# As per Wimber et al., missing responses are changed to be incorrect
d$Y[ is.na(d$Y) ] = 0
# Image type (0 = target, 1 = competitor)
d$IT = AD$ImageType - 1
# Selective retrieval ( 1 = yes, 0 = no )
d$SR = 1 - AD$Baseline
# Interaction between selective retrieval and image type
d$ITxSR = d$IT * d$SR
# Subject/item factors
d$S = as.factor( AD$Subject )
d$I = as.factor( AD$ImageNum )
# Nuisance parameter for bernoulli distribution
d$Trial = rep(1,nrow(d))
```

Here is a plot of the proportion correct by condition and subject, to give some insight into the patter of data I'm trying to fit.



The original authors argue that for targets, performance in the baseline and selective retrieval conditions should not differ. In turn, performance should be significantly lower for the selective retrieval condition with competitors relative to the baseline condition. Operating under the assumption that the two baseline conditions should not differ either, this indicates that the condition means should all be equal, except for a decrement in performance for the selective retrieval condition for competitors. I therefore created a set of covariates singling out each condition individually.

```
# Targets in the selective retrieval condition
d$TSR = numeric( nrow(d) )
d$TSR[ d$IT == 0 & d$SR == 1 ] = 1
# Competitors in the selective retrieval condition
d$CSR = numeric( nrow(d) )
d$CSR[ d$IT == 1 & d$SR == 1 ] = 1
# Targets in the baseline condition
d$TB = numeric( nrow(d) )
d$TB[ d$IT == 0 & d$SR == 0 ] = 1
# Competitors in the baseline condition
d$CB = numeric( nrow(d) )
d$CB[ d$IT == 0 & d$SR == 0 ] = 1
```

Some questions on the appropriateness of this particular comparison could be raised. For instance, why should the target and baseline conditions all be equal? The target images were practiced more often - it seems reasonable to expect improved performance. One possibility is fatigue. Since subjects saw the target images last, fatigue could mask any improvement in performance.

A main part of this exploration was to try to find additional predictors using other stages from the task. Creating these covariates is not trivial, admittably. First, we wanted to create predictors based on the performance of subjects during the training stage (in which subjects carried out a recognition memory test with feedback twice for the target images, once for the competitor images). Certain decisions needed to be made here.

For instance, a normalized score representing a subject's average performance across all training trials could be used. In contrast, a score representing a subject's performance on individual images could be used. Furthermore, should performance be weighted equally for targets and competitors, should competitors be downweighted to reflect the single training cycle, or should separate variables be used for targets and competitors?

The covariates I've created are

- TS_1: Overall performance across targets/competitors and subjects
- TS_2: Overall performance across 4 conditions and subjects
- TS_3: Overall performance across 4 conditions weighted by image performance

I provide a brief table showing the possible values that each of these covariates can take on across the 4 conditions.

```
##    IT SR TS_1 TS_2 TS_3
## 73  0  1 1.29 1.24 1.24
## 74  0  0 1.29 1.18 1.18
## 77  0  1 1.29 1.24 1.24
## 78  0  0 1.29 1.18 0.79
```

```
##    IT SR TS_1 TS_2 TS_3
## 24  1  0 0.77 0.46 0.15
## 25  1  0 0.77 0.46 0.46
## 26  1  1 0.77 0.82 0.82
## 27  1  1 0.77 0.82 0.82
```

Additionally, it would be nice to incorporate predictors based on performance during the selective retrieval stage. This becomes particularly useful since Wimber et al. report a significant correlation between the number of intrusions a subject experienced and below-baseline forgetting. Put another way, we could potentially determine whether having a higher number of intrustions predicted less of an RIF effect (which I assume is what the original authors would predict).

Again, several decisions must be made regarding the nature of the predictors. Furthermore, I've yet to properly extract the data from the selective retrieval stage from the raw output files provided by Wimber et al. The summary statistics are close (within 3 percentage points), but do not exactly match.

The covariates I've created are

- SR_T: Overall proportion of target identifications from selective retrieval
- SR_T2: Overall target identifications, but the coefficient is active only for the selective retrieval condition for competitor images.
- SR_I: Overall proportion of intrusions from selective retrieval
- SR_I2: Overall intrusions, but the coefficient is active only for the selective retrieval condition for competitor images.

I provide a brief table showing the possible values that each of these covariates can take on across the 4 conditions.

```
##    IT SR SR_T SR_T2 SR_I SR_I2
## 73  0  1 0.94     0 0.41     0
## 74  0  0 0.94     0 0.41     0
## 77  0  1 0.94     0 0.41     0
## 78  0  0 0.94     0 0.41     0
```

```
##    IT SR SR_T SR_T2 SR_I SR_I2
## 24  1  0 0.94  0.00 0.41  0.00
## 25  1  0 0.94  0.00 0.41  0.00
## 26  1  1 0.94  0.94 0.41  0.41
## 27  1  1 0.94  0.94 0.41  0.41
```

If subjects suppress the competitor images when retrieving the target images during selective retrieval, then a higher proportion of target recalls should predict worse performance in the selective retrieval condition for competitors. Conversely, a greater number of intrustions should predict better performance in this condition, because it indicates the subject failed to suppress the competitor images as well.

```
# RIF model
mRIF = glmer( Y ~ CSR + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mRIF )$coefficients[1:2,c(1,2,4)], 3 ) )
```

```
##             Estimate Std. Error Pr(>|z|)
## (Intercept)    1.516      0.114    0.000
## CSR           -0.279      0.087    0.001
```

```
# RIF model modulated by proportion of targets recalled
# during the selective retrieval stage
mRIF_v2 = glmer( Y ~ CSR + SR_T2 + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mRIF_v2 )$coefficients[1:3,c(1,2,4)], 3 ) )
```

```
##             Estimate Std. Error Pr(>|z|)
## (Intercept)    1.508      0.107    0.000
## CSR           -1.262      0.477    0.008
## SR_T2          1.203      0.574    0.036
```

```
# RIF model modulated by proportion of intrusions
# during the selective retrieval stage
mRIF_v3 = glmer( Y ~ CSR + SR_I2 + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mRIF_v3 )$coefficients[1:3,c(1,2,4)], 3 ) )
```

```
##               Estimate Std. Error Pr(>|z|)
## (Intercept)    1.514      0.112    0.000
## CSR           -0.110      0.163    0.499
## SR_I2         -0.385      0.315    0.222
```

```
# Model comparison
comp = anova( mRIF, mRIF_v2, mRIF_v3 )
# Second model is preferred
print( round( AkaikeWeights( comp, rownames(comp) )[[2]], 2 ) )
```

```
##    mRIF mRIF_v2 mRIF_v3
##    0.20    0.64    0.16
```

The second model was 3 times more likely to predict new data, according to the Akaike weights. This models assumes that performance in the baseline conditions and the selective retrieval condition for targets was equivalent. However, performance for the selective retrieval condition for competitors could be modulated by two factors. There was an overall decrement to performance, which could be further adjusted based on a normalized proportion of target recall during the selective retrieval stage. The performance of the subject who recalled the most targets was fully adjusted by the coefficient, whereas the subject with the worse recall was impacted the least by the coefficient.

The positive (significant) coefficient indicates that the decrement in performance for the selective retrieval condition was sizably decreased by better recall of targets. This seems confusing to me, as it implies that people who were better at suppressing competitors were subsequently less likely to forget these self-same competitors. This result therefore warrants further checking to make sure I didn't make a mistake.

This is complicated by feedback. I can easily imagine a scenario in which subjects experienced intrusions, saw the feedback for the correct category, and imagined a protypical example of the category which then sufficiently matched for the measure of the patterns in the BOLD response specific to the particular target image.

We can also see how the preferred RIF model compares to a suite of other models (a null model, the original model fit by the authors, and a standard interaction model).

```
# Null model
mNull = glmer( Y ~ (1|S) + (1|I), data = d,
               family = binomial(link='logit'), weights = Trial )

# Original model
mOrig = glmer( Y ~ IT + SR + ITxSR + (1|S), data = d,
               family = binomial(link='logit'), weights = Trial )
print( round( summary( mOrig )$coefficients[1:4,c(1,2,4)], 3 ) )
```

```
##               Estimate Std. Error Pr(>|z|)
## (Intercept)    1.403      0.143    0.000
## IT             0.169      0.175    0.335
## SR            -0.062      0.139    0.654
## ITxSR         -0.369      0.199    0.063
```

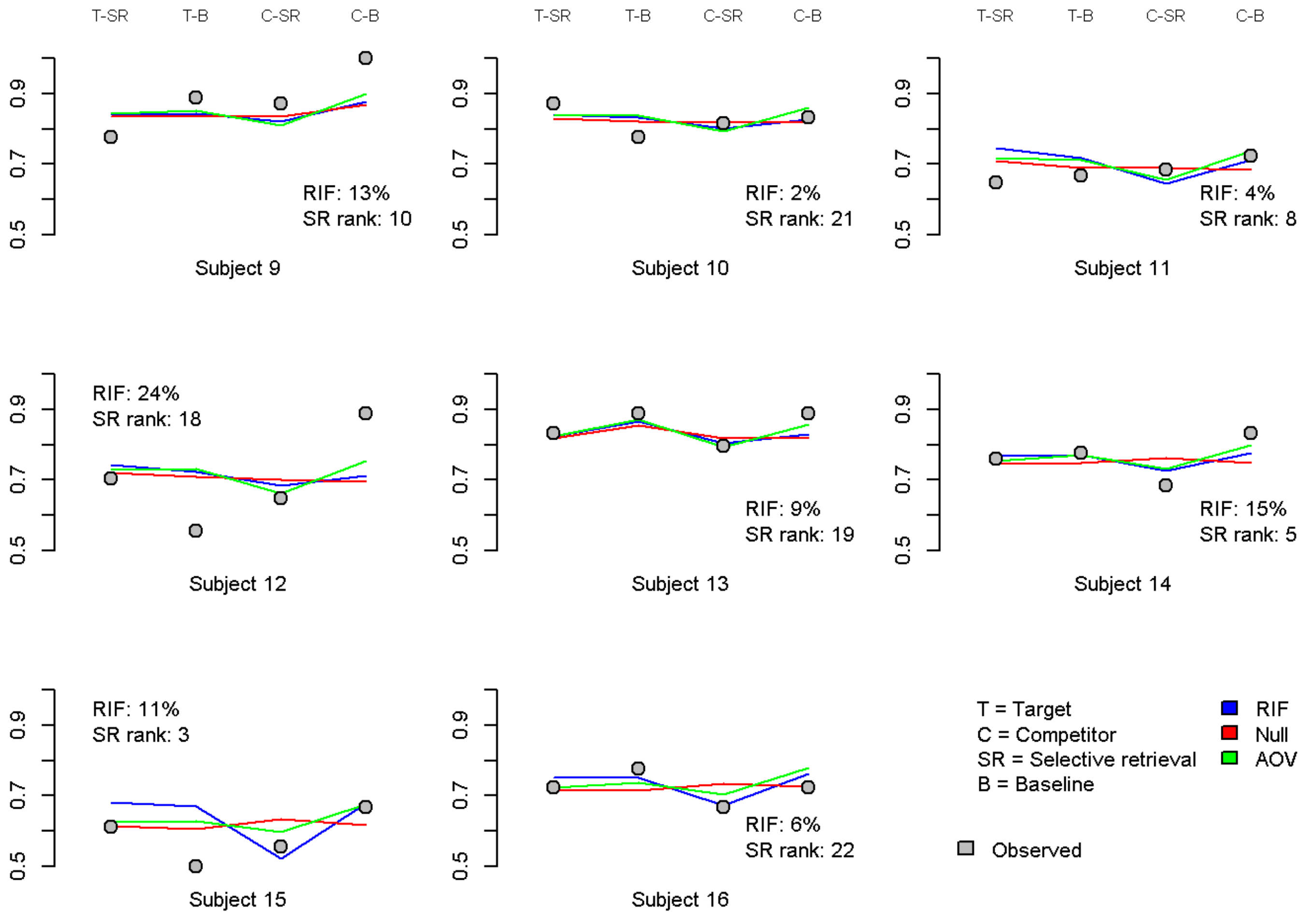
```
# Item effects
mAOV = glmer( Y ~ IT + SR + ITxSR + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mAOV )$coefficients[1:4,], 3 ) )
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.528      0.159   9.608    0.000
## IT             0.177      0.179   0.990    0.322
## SR            -0.073      0.142  -0.514    0.607
## ITxSR         -0.394      0.203  -1.936    0.053
```

```
# Model comparison
comp = anova( mNull, mOrig, mAOV, mRIF_v2 )
# The RIF model is strongly preferred
print( round( AkaikeWeights( comp, rownames(comp) )[[2]], 2 ) )
```

```
##   mNull   mOrig mRIF_v2    mAOV
##    0.00    0.00    0.84    0.15
```

The RIF model is strongly preferred. To answer why, we can look at the model predictions of performance on a subject-to-subject basis. I present a plot of 8 subjects, showing the observed proportions as dots, and the model predictions for the RIF model, null model, and the interaction model as blue, red, and green lines respectively.



This gives us some insight into how the models are doing. In particular, the null model fails because it can't capture the the shifts for the competitive images from the selective retrieval to the baseline condition. There appears to be fairly reliable dip between these conditions. In contrast, while there can be quite large shifts outside the scope of the model predictions for the remaining conditions, the shifts can be either positive or negative depending on the subject, thereby canceling each other out. The interaction model can capture this pattern, but it does so with more parameters than the RIF model, hence the preference.

We can also incorporate subject's previous overall performance for targets and for competitors. The covariate for training performance has been centered and scaled. We'll examine a model predicting performance just based on previous training, and a model that incorporates an additional decrement to competitor images in the selective retrieval condition based on the overall number of intrusions.

```
mPrac = glmer( Y ~ TS_3 + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mPrac )$coefficients[1:2,c(1,2,4)], 3 ) )
```

```
##             Estimate Std. Error Pr(>|z|)
## (Intercept)    1.387      0.097    0.000
## TS_3           0.159      0.072    0.027
```

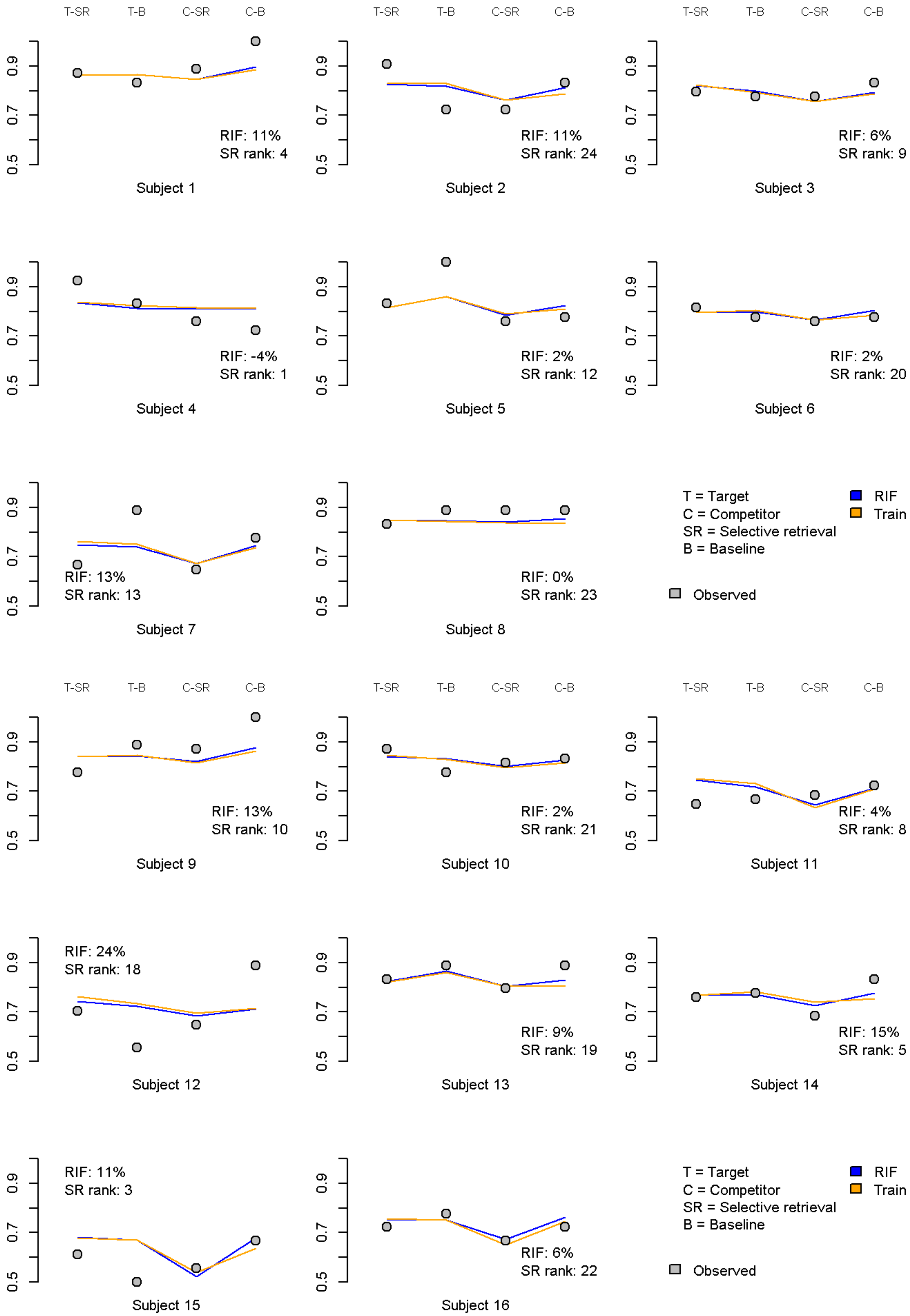
```
mPrac_v2 = glmer( Y ~ TS_3 + CSR + SR_T2 + (1|S) + (1|I), data = d,
              family = binomial(link='logit'), weights = Trial )
print( round( summary( mPrac_v2 )$coefficients[1:4,c(1,2,4)], 3 ) )
```

```
##             Estimate Std. Error Pr(>|z|)
## (Intercept)    1.475      0.100    0.000
## TS_3           0.129      0.072    0.072
## CSR           -1.339      0.461    0.004
## SR_T2          1.351      0.556    0.015
```

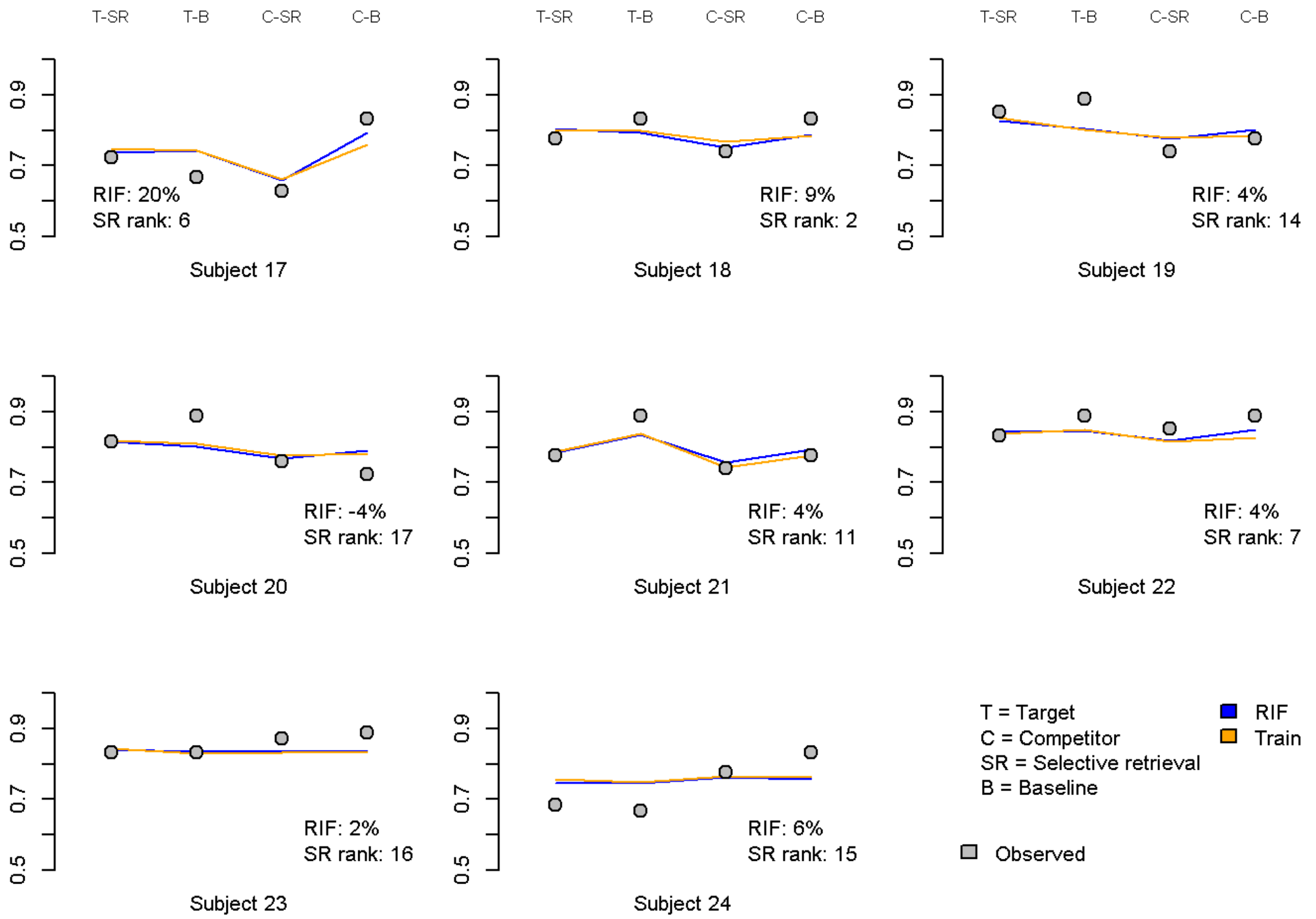
```
# Model comparison
comp = anova( mPrac, mPrac_v2, mRIF_v2 )
# The RIF model with practice is strongly preferred
print( round( AkaikeWeights( comp, rownames(comp) )[[2]], 2 ) )
```

```
##    mPrac  mRIF_v2 mPrac_v2
##     0.01     0.39     0.60
```

Previous performance weakly predicts subsequent performance. The model with the addition of the RIF effect is strongly preferred; previous overall training cannot fully account for the observed data. I present the full set of fits for each subjects for the preferred RIF model and the RIF model with practice effects.



T-SR
T-B
C-SR
C-B
0.5
0.7
0.9
RIF: 11%
SR rank: 4
Subject 1
RIF: 11%
SR rank: 24
Subject 2
RIF: 6%
SR rank: 9
Subject 3
RIF: -4%
SR rank: 1
Subject 4
RIF: 2%
SR rank: 12
Subject 5
RIF: 2%
SR rank: 20
Subject 6
RIF: 13%
SR rank: 13
Subject 7
RIF: 0%
SR rank: 23
Subject 8
T = Target
C = Competitor
SR = Selective retrieval
B = Baseline
RIF
Train
Observed
RIF: 13%
SR rank: 10
Subject 9
RIF: 2%
SR rank: 21
Subject 10
RIF: 4%
SR rank: 8
Subject 11
RIF: 24%
SR rank: 18
Subject 12
RIF: 9%
SR rank: 19
Subject 13
RIF: 15%
SR rank: 5
Subject 14
RIF: 11%
SR rank: 3
Subject 15
RIF: 6%
SR rank: 22
Subject 16



In conclusion, there appears to be a reliable dip in performance for the selective retrieval condition with competitors. However, this dip (i.e. forgetting of competitors) is strangely decreased by better recall of targets. Furthermore, examinination of the individual data indicates a several response patterns the model cannot handle. The model results reflect instead the patterns that aren't canceled out by noise. Once concern is that the bernoulli likelihood being used to model the data is somewhat inadequate, in that it does not fully capture the noisiness of the data (perhaps due to violations of the assumption of independence). I will need to check this concern more closely as well.