Memory for Emails: Analysis

Abhilasha Kumar

April 21, 2018

Reading the Data File

```
> cell_demo = read.csv("cell_demo.csv", header = TRUE, sep = ",")
> cell = read.csv("cell_withitems_complete_new.csv", header = TRUE, sep = ",")
> cell = merge(cell, cell_demo, by = "ID")
> cell$ID = as.factor(as.character(cell$ID))
> ## We need to account for the fact that contacts chosen from the end of
> ## the year will have fewer messages since it is the last time they were
> ## contacted -- so only messages from that month. We scale the number of messages by t
>
> cell$ScaledMessages = cell$Messages*cell$Month
```

0.1 Adding New Predictors

Within-Person and Grand Mean Centering

```
> library(dplyr)
> ## aggregate per subject all IVs and DVs
> cell_agg = group_by(cell, ID) %>%
+   summarise_at(vars(Accuracy, TimeJudgmentDistance,
+                     ScaledMessages, Vividness, Sentiment), mean)
> colnames(cell_agg) = c("ID", "acc_mean", "time_mean",
+                        "messages_mean", "vividness_mean", "sent_mean")
> ## merge aggregate info with long data
> cell = merge(cell, cell_agg, by = "ID", all.x = T)
> ## person and grand-mean centered scores using original and aggregate
> library(dplyr)
> cell = cell %>% mutate(acc_pc = Accuracy - acc_mean,
+                  time_pc = TimeJudgmentDistance - time_mean,
+                  messages_pc = ScaledMessages - messages_mean,
+                  vividness_pc = Vividness - vividness_mean,
+                  sent_pc = Sentiment - sent_mean)
```

1 Recipient Name Accuracy Models

Adding Time and Messages

```
> cell$scaledDays = cell$Days/100
> cell$scaled_m_pc = (cell$messages_pc/100)+1
> cell$scaled_m_mean= (cell$messages_mean/10)
> library(lme4)
> acc_pred_3_1 = glmer(data = cell, Accuracy ~ scaledDays + scaled_m_mean +
+                                 scaled_m_pc +
+                      (1|ID) + (1|ItemNo), family = "binomial")
> summary(acc_pred_3_1)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: Accuracy ~ scaledDays + scaled_m_mean + scaled_m_pc + (1 | ID) +
    (1 | ItemNo)
   Data: cell

     AIC      BIC   logLik deviance df.resid
   859.2    887.6   -423.6    847.2      841

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.2423  0.2156  0.3927  0.5392  1.3425

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 0.34154  0.5844
 ItemNo (Intercept) 0.06803  0.2608
Number of obs: 847, groups:  ID, 44; ItemNo, 31

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.06784    0.42576   2.508   0.0121 *
scaledDays   -0.65707    0.09207  -7.137 9.54e-13 ***
scaled_m_mean  0.11928    0.07543   1.581   0.1138
scaled_m_pc    0.83120    0.26188   3.174   0.0015 **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Correlation of Fixed Effects:
            (Intr) scldDy scld_m_m
scaledDays  -0.132
scaled_m_mn -0.747 -0.068
scaled_m_pc -0.550 -0.363  0.141
```

```
> fixed_eff = fixef(acc_pred_3_1)
> odds = exp(fixed_eff)
> #x = confint(acc_pred_3_1)
>
> #exp(x)
> ## CONFINT
>
> #   2.5 %      97.5 %
> # .sig01          0.32391683   0.8830534
> # .sig02          0.00000000   0.5902811
> # (Intercept)     0.19894118   1.9030035
> # scaledDays     -0.84330277  -0.4790048
> # scaled_m_mean  -0.02834944   0.2772891
> # scaled_m_pc     0.34950065   1.3881071
>
> acc_pred_4 = glmer(data = cell, Accuracy ~ scaledDays + scaled_m_mean +
+                      scaled_m_pc + scaled_m_mean:scaled_m_pc +
+                      (1|ID) + (1|ItemNo), family = "binomial",
+                     control=glmerControl(optimizer="bobyqa",
+                     optCtrl=list(maxfun=1000000)))
> summary(acc_pred_4)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula:
Accuracy ~ scaledDays + scaled_m_mean + scaled_m_pc + scaled_m_mean:scaled_m_pc +
    (1 | ID) + (1 | ItemNo)
   Data: cell
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+06))

     AIC      BIC   logLik deviance df.resid
   858.1    891.3   -422.1    844.1      840

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4395  0.2056  0.3904  0.5347  1.3810

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 0.34012  0.5832
 ItemNo (Intercept) 0.06222  0.2494
Number of obs: 847, groups:  ID, 44; ItemNo, 31

Fixed effects:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            0.11872    0.69802   0.170  0.86495
```

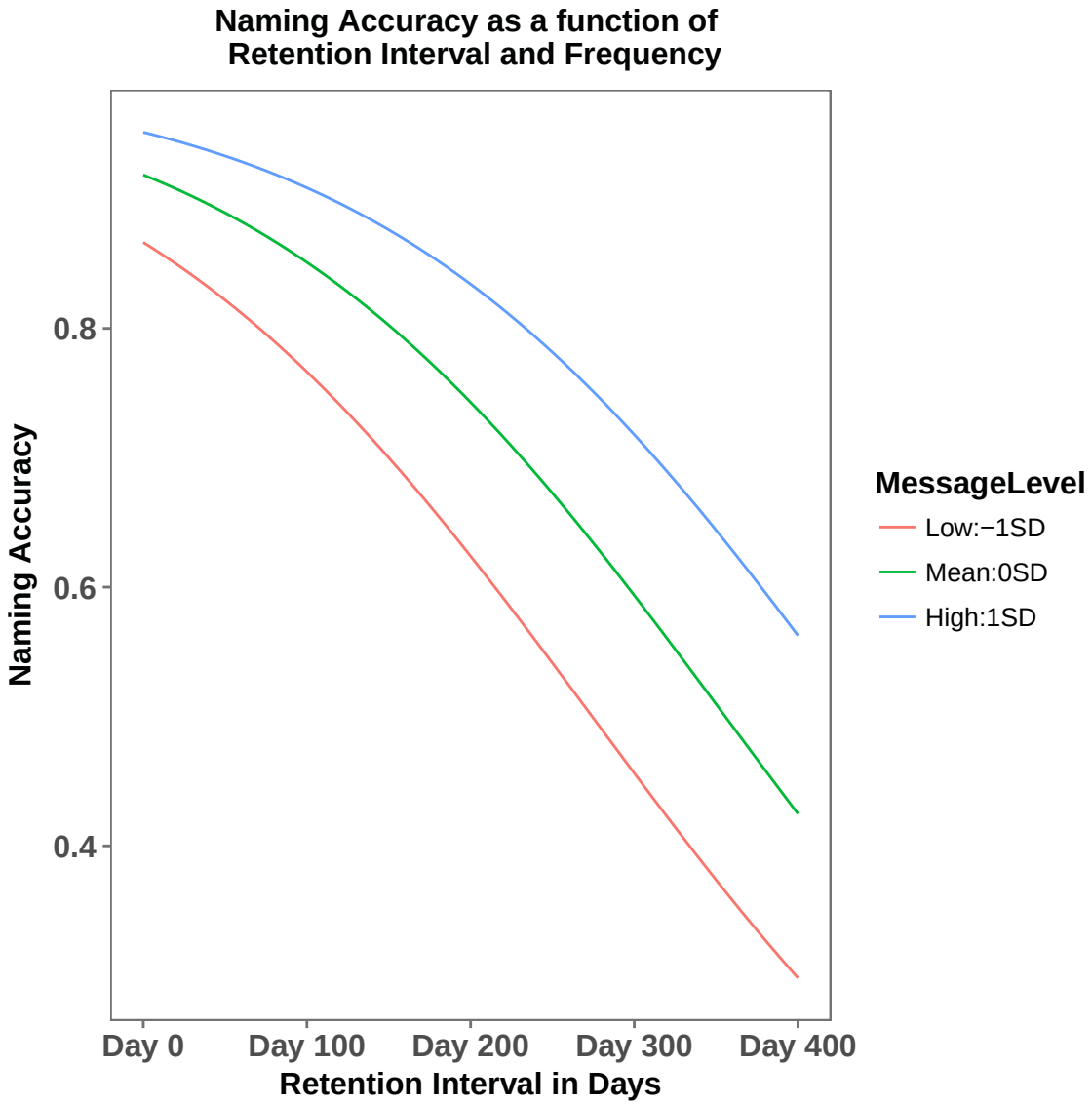
```
scaledDays                 -0.68195    0.09373  -7.275 3.45e-13 ***
scaled_m_mean               0.28575    0.12296   2.324  0.02013 *
scaled_m_pc                 1.84082    0.64490   2.854  0.00431 **
scaled_m_mean:scaled_m_pc -0.17039    0.09330  -1.826  0.06781 .
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Correlation of Fixed Effects:
            (Intr) scldDy scld_m_m scld_m_p
scaledDays   0.066
scaled_m_mn -0.903 -0.186
scaled_m_pc -0.861 -0.314  0.739
scld_m_m:__  0.822  0.194 -0.794    -0.924
```

```
> anova(acc_pred_3_1, acc_pred_4) ## only marginally better, hence choosing simpler mode
```

```
Data: cell
Models:
acc_pred_3_1: Accuracy ~ scaledDays + scaled_m_mean + scaled_m_pc + (1 | ID) +
acc_pred_3_1:     (1 | ItemNo)
acc_pred_4: Accuracy ~ scaledDays + scaled_m_mean + scaled_m_pc + scaled_m_mean:scaled_m
acc_pred_4:     (1 | ID) + (1 | ItemNo)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
acc_pred_3_1  6 859.17 887.62 -423.59   847.17
acc_pred_4    7 858.11 891.30 -422.05   844.11 3.0654      1    0.07998 .
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

2 Plotting the Model for Naming Accuracy

The final model for accuracy is accpred3, with time and messagespc and messagesmean. We are going to try and plot the fitted values from this model below:



3 Model for Temporal Dating Error

```
> ## removing trials in which participant did not provide time estimate
>
> cell_td = cell %>% filter(GuessedMonth != "-1")
> library(optimx)
```

```
> library(lme4)
> library(lmerTest)
> time_model_0 = lmer(data = cell_td, TimeJudgmentDistance ~ scaledDays +
```

```
+                         (1|ID),
+       control = lmerControl(optimizer = "optimx", calc.derivs = FALSE,
+       optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
> summary(time_model_0)
```

```
Linear mixed model fit by REML t-tests use Satterthwaite approximations to
  degrees of freedom [lmerMod]
Formula: TimeJudgmentDistance ~ scaledDays + (1 | ID)
   Data: cell_td
Control:
lmerControl(optimizer = "optimx", calc.derivs = FALSE, optCtrl = list(method = "nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 2690.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1935 -0.6071 -0.1629  0.1878  7.2854

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.000    0.000
 Residual             1.992    1.411
Number of obs: 761, groups:  ID, 44

Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  0.71604    0.08832 759.00000   8.107 2.22e-15 ***
scaledDays   0.27047    0.04826 759.00000   5.605 2.92e-08 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Correlation of Fixed Effects:
           (Intr)
scaledDays -0.815
```

```
> time_model_1 = lmer(data = cell_td, TimeJudgmentDistance ~ scaledDays +
+                         (scaledDays|ID),
+       control = lmerControl(optimizer = "optimx", calc.derivs = FALSE,
+       optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
> summary(time_model_1)
```

```
Linear mixed model fit by REML t-tests use Satterthwaite approximations to
  degrees of freedom [lmerMod]
Formula: TimeJudgmentDistance ~ scaledDays + (scaledDays | ID)
   Data: cell_td
Control:
lmerControl(optimizer = "optimx", calc.derivs = FALSE, optCtrl = list(method = "nlminb",
```

```
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 2681.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7648 -0.6456 -0.1259  0.1887  7.3696

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ID       (Intercept) 0.03605  0.1899
          scaledDays  0.05940  0.2437   -1.00
 Residual             1.90315  1.3795
Number of obs: 761, groups:  ID, 44

Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  0.69398    0.09160 122.23000   7.576 7.58e-12 ***
scaledDays   0.29799    0.06137  42.99000   4.856 1.62e-05 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Correlation of Fixed Effects:
           (Intr)
scaledDays -0.802
```

```
> anova(time_model_0, time_model_1) ### model with random slope better
```

```
Data: cell_td
Models:
object: TimeJudgmentDistance ~ scaledDays + (1 | ID)
..1: TimeJudgmentDistance ~ scaledDays + (scaledDays | ID)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
object  4 2689.9 2708.4 -1340.9   2681.9
..1     6 2685.9 2713.7 -1336.9   2673.9 7.9873      2    0.01843 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

```
> cell_td = cell_td %>%
+   mutate(DatingErrorType = ifelse(DatingError >0 , "Telescoping", "Time Expansion"))
> cell_td$DatingErrorType = as.factor(as.character(cell_td$DatingErrorType))
> time_model_2 = lmer(data = cell_td, TimeJudgmentDistance ~
+                       scaledDays*DatingErrorType +
+                       (scaledDays|ID) + (1|ItemNo))
> anova(time_model_1, time_model_2) ## model with telescoping better
```

```
Data: cell_td
Models:
```

```
object: TimeJudgmentDistance ~ scaledDays + (scaledDays | ID)
..1: TimeJudgmentDistance ~ scaledDays * DatingErrorType + (scaledDays |
..1:     ID) + (1 | ItemNo)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
object  6 2685.9 2713.7 -1336.9   2673.9
..1     9 2547.2 2589.0 -1264.6   2529.2 144.64      3  < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

```
> #confint(time_model_2)
>
> time_model_3 = lmer(data = cell_td, TimeJudgmentDistance ~
+                       scaledDays*DatingErrorType*Sentiment +
+                        (scaledDays|ID) + (1|ItemNo))
> summary(time_model_3)
```

```
Linear mixed model fit by REML t-tests use Satterthwaite approximations to
  degrees of freedom [lmerMod]
Formula: TimeJudgmentDistance ~ scaledDays * DatingErrorType * Sentiment +
    (scaledDays | ID) + (1 | ItemNo)
   Data: cell_td

REML criterion at convergence: 2561.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6092 -0.4833 -0.3076  0.1695  8.3608

Random effects:
 Groups   Name        Variance  Std.Dev.  Corr
 ID       (Intercept) 3.474e-02 1.864e-01
          scaledDays  5.235e-02 2.288e-01 -1.00
 ItemNo   (Intercept) 8.605e-15 9.276e-08
 Residual             1.554e+00 1.247e+00
Number of obs: 761, groups:  ID, 44; ItemNo, 31

Fixed effects:
                                                    Estimate Std. Error
(Intercept)                                         0.929382   0.133486
scaledDays                                          0.432894   0.079587
DatingErrorTypeTime Expansion                      -0.523482   0.184963
Sentiment                                          -0.006666   0.010038
scaledDays:DatingErrorTypeTime Expansion           -0.380036   0.105495
scaledDays:Sentiment                                0.010077   0.006053
DatingErrorTypeTime Expansion:Sentiment             0.044897   0.015393
scaledDays:DatingErrorTypeTime Expansion:Sentiment -0.026659   0.009424
                                                          df t value Pr(>|t|)
(Intercept)                                       355.400000   6.962 1.62e-11
```

```
scaledDays                                          128.900000   5.439 2.59e-07
DatingErrorTypeTime Expansion                       714.100000  -2.830 0.004783
Sentiment                                           709.900000  -0.664 0.506866
scaledDays:DatingErrorTypeTime Expansion            742.200000  -3.602 0.000336
scaledDays:Sentiment                                752.700000   1.665 0.096393
DatingErrorTypeTime Expansion:Sentiment             731.900000   2.917 0.003645
scaledDays:DatingErrorTypeTime Expansion:Sentiment  733.700000  -2.829 0.004800

(Intercept)                                         ***
scaledDays                                          ***
DatingErrorTypeTime Expansion                       **
Sentiment
scaledDays:DatingErrorTypeTime Expansion            ***
scaledDays:Sentiment                                .
DatingErrorTypeTime Expansion:Sentiment             **
scaledDays:DatingErrorTypeTime Expansion:Sentiment  **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Correlation of Fixed Effects:
            (Intr) scldDy DtETTE Sntmnt scD:DETTE sclD:S DETTE:
scaledDays  -0.797
DtngErrrTTE -0.688  0.505
Sentiment   -0.501  0.387  0.359
scldD:DETTE  0.527 -0.603 -0.801 -0.287
scldDys:Snt  0.380 -0.499 -0.271 -0.767  0.371
DtngErTTE:S  0.326 -0.250 -0.524 -0.650  0.437     0.499
scD:DETTE:S -0.242  0.319  0.407  0.489 -0.543    -0.642 -0.801
```

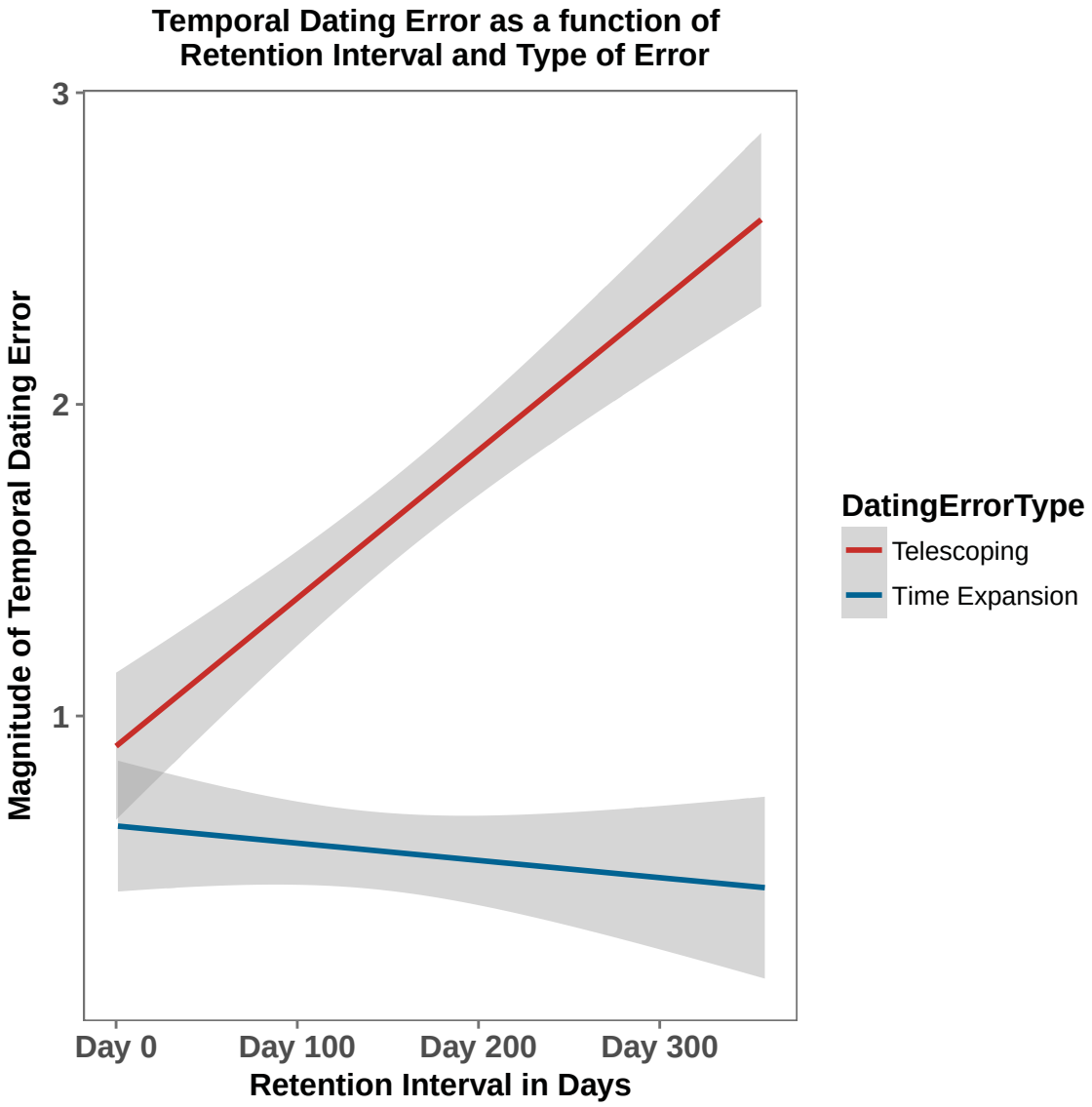
```
> anova(time_model_2, time_model_3) ## model with interaction better
```

```
Data: cell_td
Models:
object: TimeJudgmentDistance ~ scaledDays * DatingErrorType + (scaledDays |
object:     ID) + (1 | ItemNo)
..1: TimeJudgmentDistance ~ scaledDays * DatingErrorType * Sentiment +
..1:     (scaledDays | ID) + (1 | ItemNo)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
object  9 2547.2 2589.0 -1264.6   2529.2
..1    13 2540.4 2600.7 -1257.2   2514.4 14.846      4   0.005031 **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

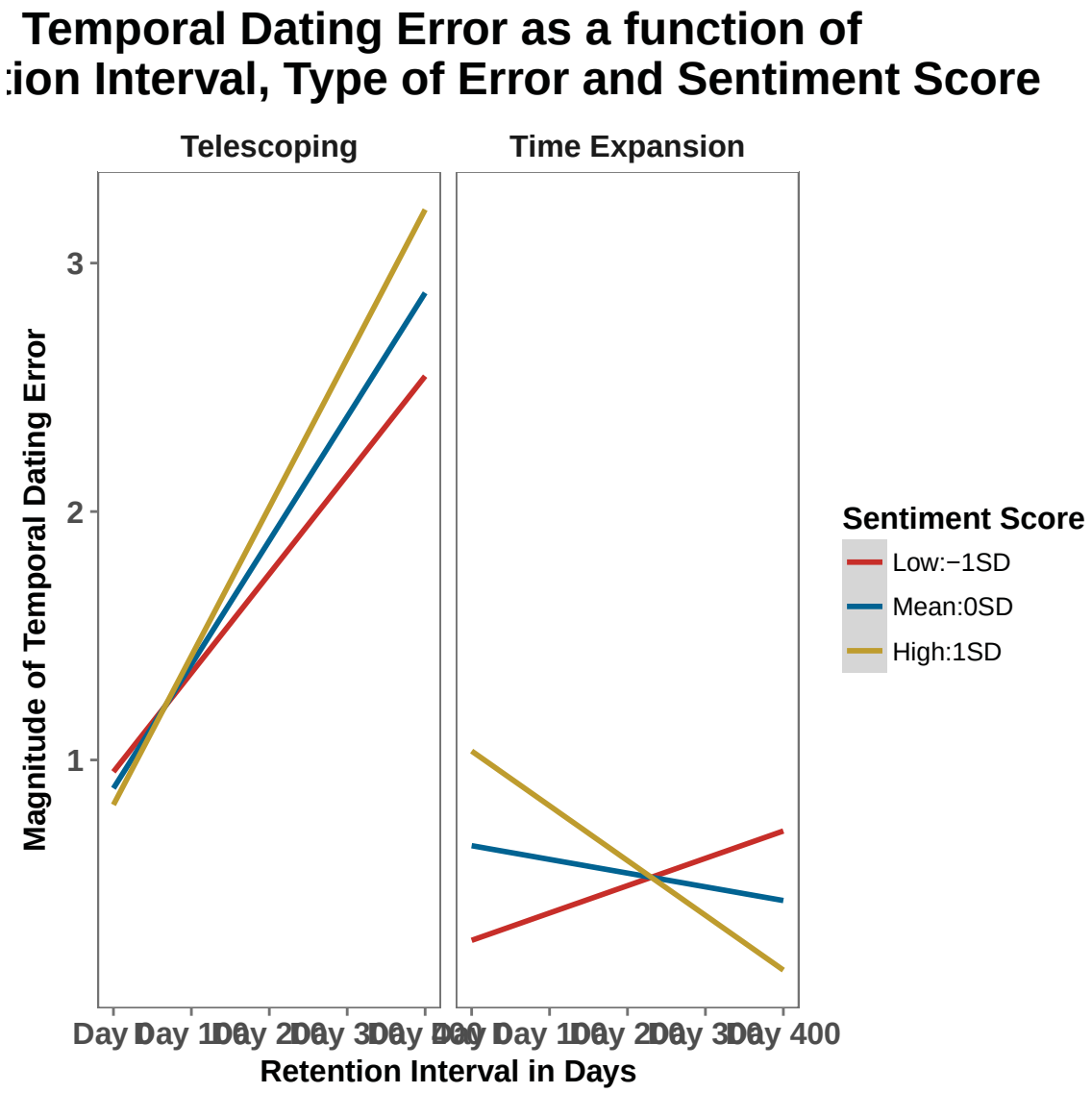
```
>
> # > confint(time_model_9)
> # Computing profile confidence intervals ...
> #                                                    2.5 %       97.5 %
> # .sig01                                        0.00000000  0.190317012
```

```
> #  .sig02                                                  -1.00850935   1.000000000
> #  .sig03                                                   0.08141825   0.250670906
> #  .sig04                                                   0.00000000           Inf
> #  .sigma                                                   1.17874882   1.308565534
> #  (Intercept)                                              0.66577143   1.191413354
> #  scaledDays                                               0.27700430   0.591695270
> #  DatingErrorTypeTime  Expansion                          -0.88469137  -0.161632813
> #  Sentiment                                               -0.02649751   0.013069709
> #  scaledDays:DatingErrorTypeTime  Expansion               -0.58669259  -0.174282554
> #  scaledDays:Sentiment                                    -0.00183579   0.022065459
> #  DatingErrorTypeTime  Expansion:Sentiment                 0.01458253   0.074900614
> #  scaledDays:DatingErrorTypeTime  Expansion:Sentiment     -0.04516634  -0.007777518
>
```

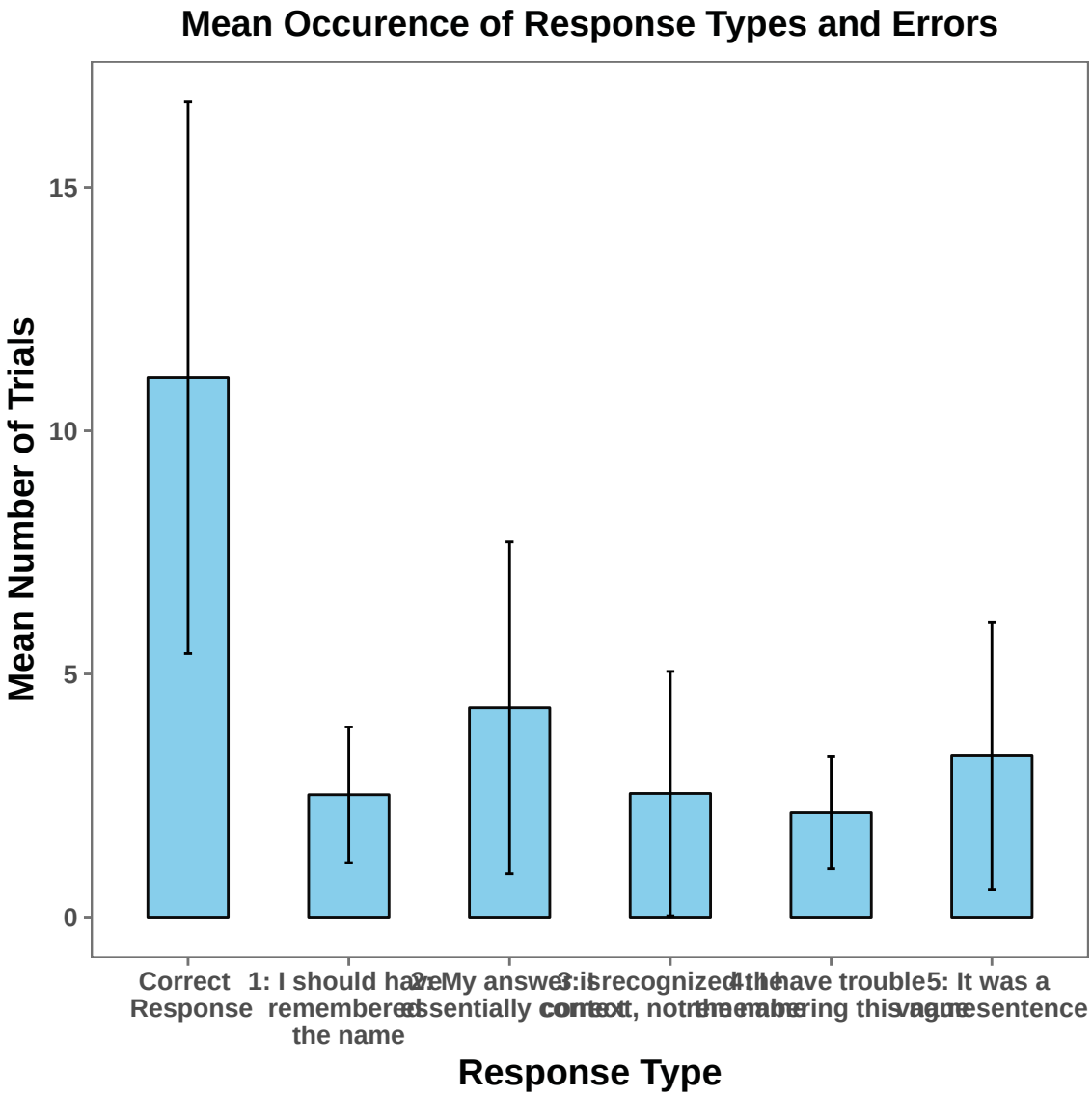
3.1 Plotting Temporal Model 2



3.2 Plotting Temporal Model 3



4 Descriptive Graphs: Error Judgments



5 Retrieval States