Mixed-effects regression and eye-tracking data

Lecture 2 of advanced regression for linguists

Martijn Wieling

Department of Humanities Computing, University of Groningen

Cambridge, January 13, 2015

This lecture

- Introduction
  - Gender processing in Dutch
  - Eye-tracking to reveal gender processing
- Design
- Analysis
- Conclusion

Gender processing in Dutch

- The goal of this study is to investigate if Dutch people use grammatical gender to anticipate upcoming words
  - This study was conducted together with Hanneke Loerts and is published in the *Journal of Psycholinguistic Research* (Loerts, Wieling and Schmid, 2012)
- What is grammatical gender?
  - Gender is a property of a noun
  - Nouns are divided into classes: masculine, feminine, neuter, ...
  - E.g., *hond* ('dog') = common, *paard* ('horse') = neuter
- The gender of a noun can be determined from the forms of other elements syntactically related to it (Matthews, 1997: 36)

Gender in Dutch

Gender	Definite article	Adjective in definite NPs	Adjective in Indefinite NPs
Common	De hond	De mooie hond	Een mooie hond
English equivalent	*The$_{com}$ dog$_{com}$*	*The$_{com}$ beautiful dog$_{com}$*	*A beautiful$_{com}$ dog$_{com}$*
Neuter	Het huis	Het mooie huis	Een mooi huis
English equivalent	*The$_{neu}$ house$_{neu}$*	*The$_{neu}$ beautiful house$_{neu}$*	*A beautiful$_{neu}$ house$_{neu}$*

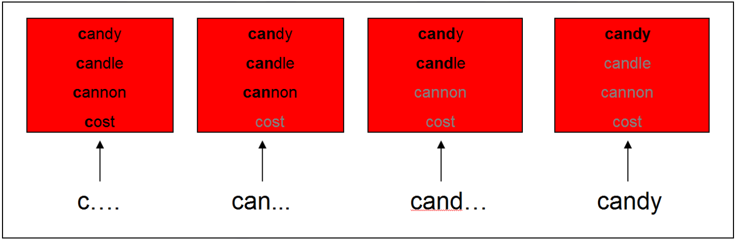
- Gender in Dutch: 70% common, 30% neuter
  - When a noun is diminutive it is always neuter
- Gender is unpredictable from the root noun and hard to learn
  - Children overgeneralize until the age of 6 (Van der Velde, 2004)

Why use eye tracking?

- Eye tracking reveals incremental processing of the listener during the time course of the speech signal
- As people tend to look at what they hear (Cooper, 1974), lexical competition can be tested

Testing lexical competition using eye tracking

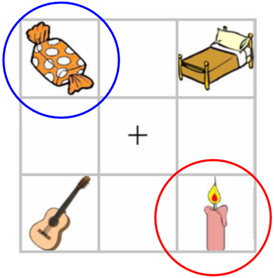
- Cohort Model (Marslen-Wilson & Welsh, 1978): Competition between words is based on word-initial activation



- This can be tested using the visual world paradigm: following eye movements while participants receive auditory input to click on one of several objects on a screen

Support for the Cohort Model

- Subjects hear: “Pick up the candy” (Tanenhaus et al., 1995)
- Fixations towards target (Candy) *and* competitor (Candle): support for the Cohort Model

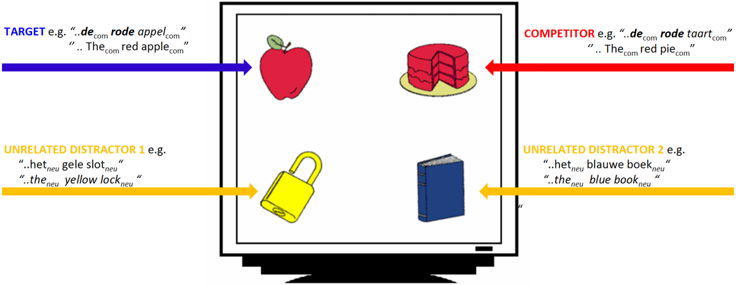


Lexical competition based on syntactic gender

- Other models of lexical processing state that lexical competition occurs based on all acoustic input (e.g., TRACE, Shortlist, NAM)
- Does gender information restrict the possible set of lexical candidates?
  - I.e. if you hear *de*, will you focus more on an image of a dog (*de hond*) than on an image of a horse (*het paard*)?
  - Previous studies (e.g., Dahan et al., 2000 for French) have indicated gender information restricts the possible set of lexical candidates
- In the following, we will investigate if this also holds for Dutch with its difficult gender system using the visual world paradigm
  - We analyze the data using mixed-effects regression in `R`

Experimental design

- 28 Dutch participants heard sentences like:
  - *Klik op de rode appel* ('click on the red apple')
  - *Klik op het plaatje met een blauw boek* ('click on the image of a blue book')
- They were shown 4 nouns varying in color and gender
  - Eye movements were tracked with a Tobii eye-tracker (E-Prime extensions)



university of groningen

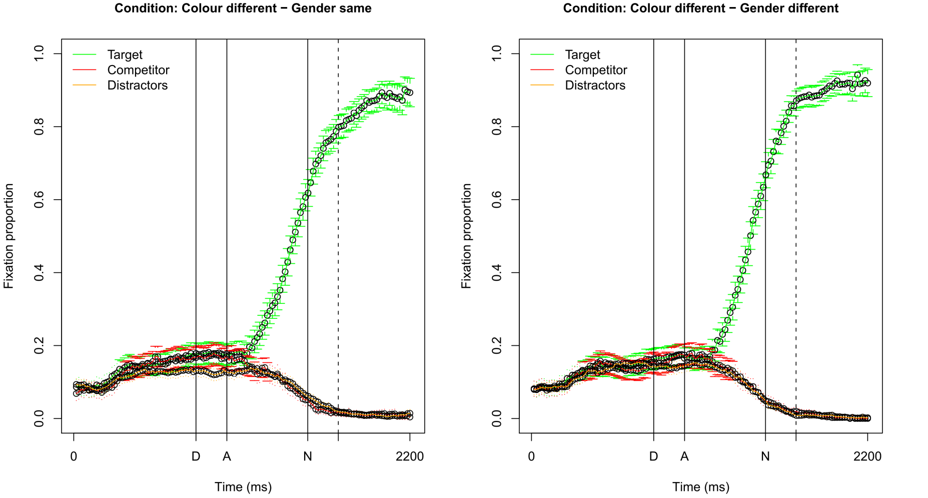
NWO
Netherlands Organisation for Scientific Research

Experimental design: conditions

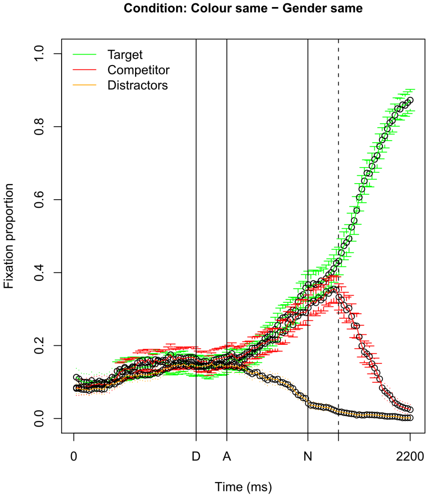
Target	Competitor	Gender Competitor	Colour Competitor
De$_{com}$ rode appel$_{com}$ *The$_{com}$ red apple$_{com}$*	Het$_{neu}$ groene bureau$_{neu}$ *The$_{neu}$ green desk$_{neu}$* 	*Different*	*Different*
	De$_{com}$ gele zon$_{com}$ *The$_{com}$ yellow sun$_{com}$* 	*Same*	*Different*
	Het$_{neu}$ rode hart$_{neu}$ *The$_{neu}$ red heart$_{neu}$* 	*Different*	*Same*
	De$_{com}$ rode taart$_{com}$ *The$_{com}$ red cake$_{com}$* 	*Same*	*Same*

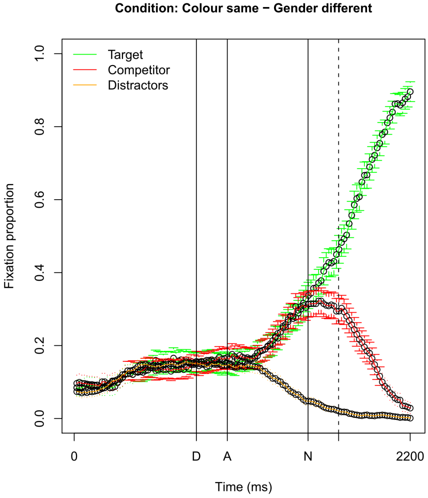
- Subjects were shown 96 different screens
  - 48 screens for indefinite sentences (*klik op het plaatje met een rode appel*)
  - 48 screens for definite sentences (*klik op de rode appel*)

Visualizing fixation proportions: different color



Visualizing fixation proportions: same color





Which dependent variable?

- Difficulty 1: choosing the dependent variable
  - Fixation difference between Target and Competitor
  - Fixation proportion on Target - requires transformation to empirical logit, to ensure the dependent variable is unbounded: $log(\frac{(y+0.5)}{(N-y+0.5)})$
  - ...
- Difficulty 2: selecting a time span
  - Note that about 200 ms. is needed to plan and launch an eye movement
  - It is possible (and better) to take every individual sampling point into account, but we will opt for the simpler approach here (in contrast to lecture 4)
- In this lecture we use:
  - The difference in fixation time between Target and Competitor
  - Averaged over the time span starting 200 ms. after the onset of the determiner and ending 200 ms. after the onset of the noun (about 800 ms.)
  - This ensures that gender information has been heard and processed, both for the definite and indefinite sentences

Independent variables

- Variable of interest
  - Competitor gender vs. target gender
- Variables which could be important
  - Competitor color vs. target color
  - Gender of target (common or neuter)
  - Definiteness of target
- Participant-related variables
  - Gender (male/female), age, education level
  - Trial number
- Design control variables
  - Competitor position vs. target position (up-down or down-up)
  - Color of target
  - ... (anything else you are not interested in, but potentially problematic)

Some remarks about data preparation

- Check if variables correlate highly
  - If so: exclude variable / combine variables (residualization is not OK: Wurm & Fisicaro, 2014)
  - See Chapter 6.2.2 of Baayen (2008)
- Check if numerical variables are normally distributed
  - If not: try to make them normal (e.g., logarithmic or inverse transformation)
  - Note that your dependent variable does not need to be normally distributed (the residuals of your model do!)
- Center your numerical predictors when doing mixed-effects regression
  - See previous lecture

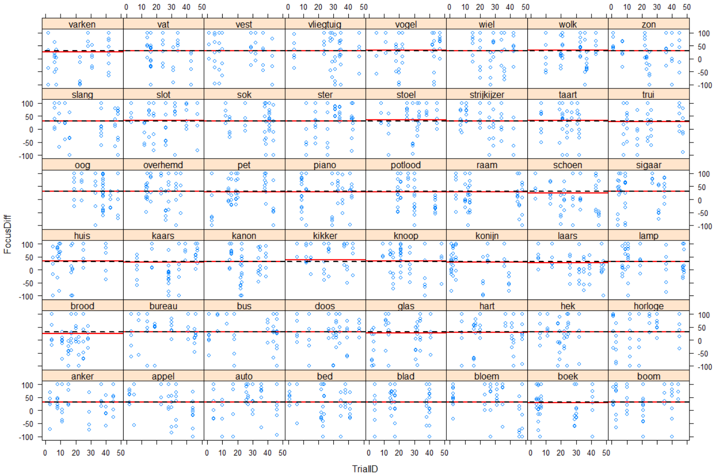
Our data

```
> head(eye)

  Subject  Item TargetDefinite TargetNeuter TargetColor TargetBrown TargetPlace
1    S300 appel              1            0         red           0           1
2    S300 appel              0            0         red           0           2
3    S300   vat              1            1       brown           1           4
4    S300   vat              0            1       brown           1           1
5    S300  boek              1            1        blue           0           4
6    S300  boek              0            1        blue           0           1
  TargetTopRight CompColor CompPlace TupCdown CupTdown TrialID Age IsMale
1              0       red         2        0        0      44  52      0
2              1     brown         4        1        0       2  52      0
3              0    yellow         2        0        1      14  52      0
4              0     brown         3        1        0      43  52      0
5              0      blue         3        0        0       5  52      0
6              0    yellow         3        1        0      30  52      0
  Edulevel SameColor SameGender TargetPerc  CompPerc  FocusDiff
1        1         1          1   40.90909  6.818182  34.090909
2        1         0          0   63.63636  0.000000  63.636364
3        1         0          0   47.72727 43.181818   4.545455
4        1         1          0   27.90698  9.302326  18.604651
5        1         1          0   11.11111 25.000000 -13.888889
6        1         0          1   23.80952 50.000000 -26.190476
```

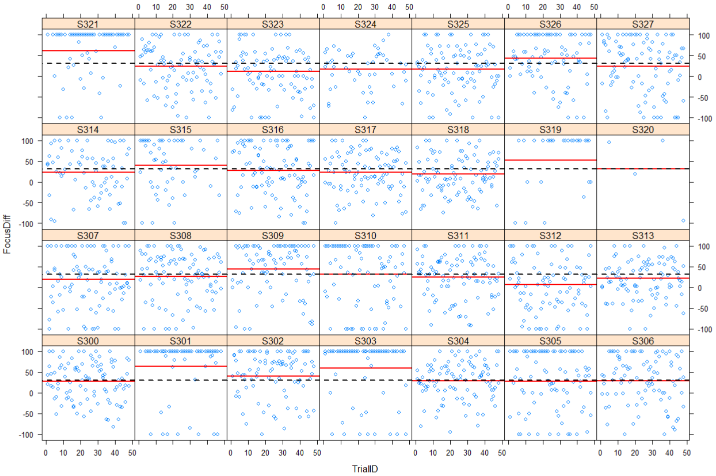
Our first mixed-effects regression model

```
# A model having only random intercepts for Subject and Item
> library(lme4) # version 1.1.7
> model = lmer( FocusDiff ~ (1|Subject) + (1|Item) , data=eye )

# Show results of the model
> summary( model )
...
Random effects:
 Groups   Name        Variance Std.Dev.
 Item     (Intercept)   33.5    5.788
 Subject  (Intercept)  294.3   17.156
 Residual             3294.0   57.393
Number of obs: 2266, groups: Item, 48; Subject, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)   28.826      3.608   7.989
...
```

By-item random intercepts



By-subject random intercepts



Is a by-item analysis necessary?

```
# comparing two models (REML=T for random effect comparison: default)
> model1 = lmer(FocusDiff ~ (1|Subject), data=eye)
> model2 = lmer(FocusDiff ~ (1|Subject) + (1|Item), data=eye)
> AIC(model1) - AIC(model2)
[1] 1.965797
```

- the AIC value is lower than 2, so we can exclude the by-item random intercept
  - This indicates that the different conditions were well-controlled in the research design

Adding a fixed-effect factor

```
# model with fixed effects, but no random-effect factor for Item
> model3 = lmer(FocusDiff ~ SameColor + (1|Subject), data=eye)
> summary(model3)
...
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept)  233.9   15.29
 Residual             2866.4   53.54
Number of obs: 2266, groups:  Subject, 28

Fixed effects:
           Estimate Std. Error t value
(Intercept)  49.655      3.331   14.91
SameColor   -43.305      2.260  -19.16
...
```

- `SameColor` is highly important as $|t| > 2$
  - Negative estimate: more difficult to distinguish target from competitor
- We need to test if the effect of `SameColor` varies per subject
  - If there is much between-subject variation, this will influence the significance of the variable in the fixed effects

Testing for a random slope

```
# as SameColor is a binary predictor (contrasting it with the intercept),
# it needs to be correlated with the random intercept
> model4 = lmer(FocusDiff ~ SameColor + (1+SameColor|Subject), data=eye)
> AIC(model3) - AIC(model4)
[1] 9.81883

> summary(model4)
...
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept)  371.6    19.28
          SameColor    143.7    11.99    -0.86
 Residual             2837.1    53.26
Number of obs: 2266, groups:  Subject, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)   49.576      4.030   12.30
SameColor    -44.197      3.234  -13.66
```

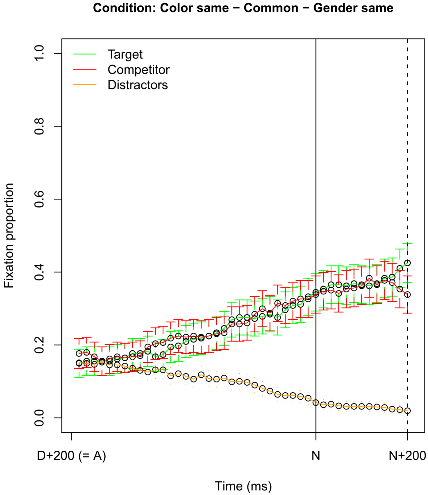
- Note `SameColor` is still highly significant as $|t| > 2$ (absolute value)

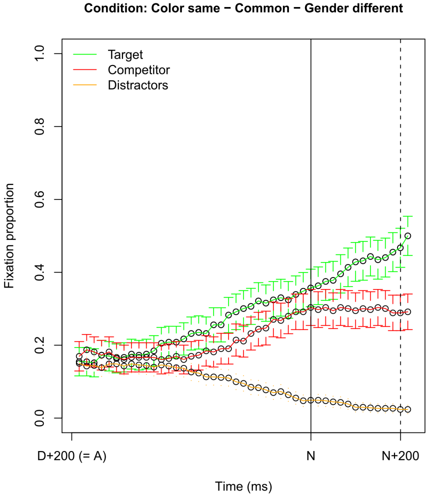
Investigating the gender effect

```
> model5 = lmer(FocusDiff ~ SameColor + SameGender +
                          (1+SameColor|Subject), data=eye)
> summary(model5)
...
Fixed effects:
            Estimate Std. Error t value
(Intercept)  49.015      4.177  11.735
SameColor   -44.212      3.237 -13.657
SameGender    1.149      2.241   0.513
...
```

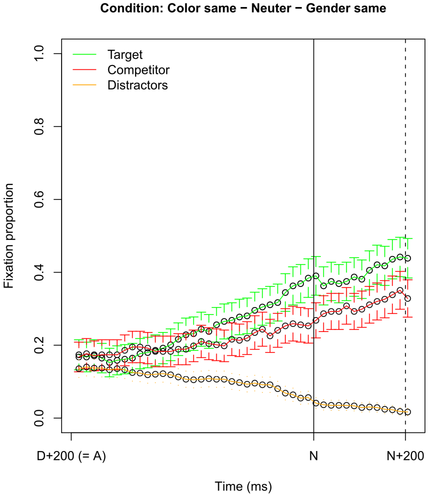
- It seems there is no gender effect...
- Perhaps there is an effect of common vs. neuter gender?

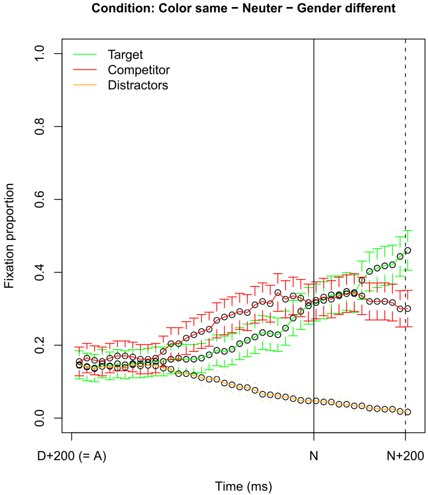
Visualizing fixation proportions: target common





Visualizing fixation proportions: target neuter





Testing the interaction

```
> model6 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                           (1+SameColor|Subject), data=eye)
> summary(model6)
...
Fixed effects:
                        Estimate Std. Error t value
(Intercept)               54.342      4.470  12.158
SameColor                -44.323      3.232 -13.712
SameGender                -5.918      3.144  -1.882
TargetNeuter             -10.650      3.153  -3.377
SameGender:TargetNeuter   14.286      4.469   3.197
...
```

- There is clear support for an interaction (the *t*'s are all close to 2)
- These results are in line with the previous fixation proportion graphs

Testing if the interaction yields an improved model

```
# To compare models differing in fixed effects, we specify REML=F.
# We compare to the best model we had before, and include TargetNeuter as
# it is also significant by itself.

> model6a = lmer(FocusDiff ~ SameColor + SameGender + TargetNeuter +
                             (1+SameColor|Subject), data=eye, REML=F)
> model6b = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                             (1+SameColor|Subject), data=eye, REML=F)
> AIC(model6a) - AIC(model6b)
[1] 8.211918
```

- The interaction improves the model significantly
  - Unfortunately, we do not have an explanation for the strange neuter pattern
- Note that we still need to test the variables for inclusion as random slopes (we do this in the lab session)

Adding a factor to the model

```
# set a reference level for the factor
> eye$TargetColor = relevel( eye$TargetColor, "brown" )
> model7 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                          TargetColor + (1+SameColor|Subject), data=eye)
> summary(model7)
...
Fixed effects:
                        Estimate Std. Error t value
(Intercept)               41.364      5.097   8.116
SameColor                -44.464      3.208 -13.861
SameGender                -5.778      3.125  -1.849
TargetNeuter             -10.799      3.134  -3.445
TargetColorblue           11.545      3.660   3.154
TargetColorgreen          16.135      3.648   4.423
TargetColorred            16.673      3.646   4.572
TargetColoryellow         18.290      3.649   5.013
SameGender:TargetNeuter   14.227      4.442   3.203
...
```

Comparing different factor levels

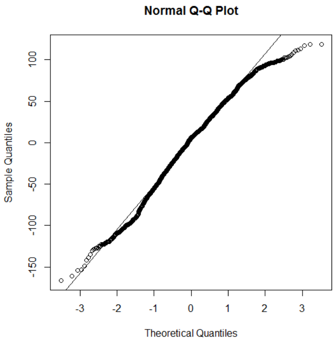
```
> library(multcomp)
> summary(glht(model7,linfct=mcp(TargetColor = "Tukey")))

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lmer(formula = FocusDiff ~ SameColor + SameGender * TargetNeuter +
TargetColor + (1 + SameColor | Subject), data = eye)

Linear Hypotheses:
                     Estimate Std. Error z value Pr(>|z|)
blue - brown == 0     11.5449     3.6601   3.154    0.014 *
green - brown == 0    16.1354     3.6478   4.423   <0.001 ***
red - brown == 0      16.6728     3.6465   4.572   <0.001 ***
yellow - brown == 0   18.2902     3.6486   5.013   <0.001 ***
green - blue == 0      4.5905     3.4466   1.332    0.671
red - blue == 0        5.1279     3.4468   1.488    0.570
yellow - blue == 0     6.7453     3.4484   1.956    0.288
red - green == 0       0.5374     3.4329   0.157    1.000
yellow - green == 0    2.1548     3.4361   0.627    0.971
yellow - red == 0      1.6174     3.4347   0.471    0.990
```

Simplifying the factor in a contrast

```
> eye$TargetBrown = (eye$TargetColor == "brown")*1
> model8 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                            TargetBrown + (1+SameColor|Subject), data=eye)
> summary(model8)
...
Fixed effects:
                        Estimate Std. Error t value
(Intercept)               57.075      4.498  12.689
SameColor                -44.500      3.216 -13.837
SameGender                -5.809      3.126  -1.858
TargetNeuter             -10.822      3.135  -3.452
TargetBrown              -15.677      2.982  -5.257
SameGender:TargetNeuter   14.259      4.443   3.210
...

# model7b and model8b: REML=F instead of the default (TRUE)
> AIC(model8b) - AIC(model7b) # N.B. model7b is more complex
[1] -1.770693
```

How well does the model fit?

```
# "explained variance" of the model (r-squared)
> cor( eye$FocusDiff , fitted( model8 ) )^2
[1] 0.2295258

> qqnorm( resid( model8 ) )
> qqline( resid( model8 ) )
```



Model criticism

```
# remove items which the model has trouble fitting
> eye2 = eye[abs(scale(resid(model8))) < 2.5,]

> 1 - (nrow(eye2) / nrow(eye))
[1] 0.00353045 # 0.35% of the data removed

> model9 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                            TargetBrown + (1+SameColor|Subject),
               data=eye2)

> cor( eye2$FocusDiff , fitted( model9 ) )^2
[1] 0.2425514 # improved explained variance (was 0.23)

> summary(model9) # all variables significant
...
Fixed effects:
                        Estimate Std. Error t value
(Intercept)               58.091      4.552  12.761
SameColor                -45.556      3.339 -13.644
SameGender                -6.321      3.086  -2.049
TargetNeuter             -10.517      3.099  -3.394
TargetBrown              -15.792      2.945  -5.362
SameGender:TargetNeuter   14.815      4.390   3.375
...
```

Many more things to do...

- We need to see if the significant fixed effects remain significant when adding these variables as random slopes per subject
- There are other variables we should test (e.g., education level)
- Model criticism should be applied *after* these steps

- We will experiment with these issues in the lab session after the break!
  - We use a subset of the data (only same color)
  - Simple `R`-functions are used to generate all plots

What you should remember...

- Mixed-effects regression offers an easy-to-use approach to obtain generalizable results even when your design is not completely balanced
- Mixed-effects regression models allow a fine-grained inspection of the variability of the random effects, which may provide additional insight in your data
- Mixed-effects regression models are easy to construct in R!
- We analyzed this data in a *non-optimal* way: rather than averaging over a timespan, it is better to predict the focus for every individual timepoint using generalized additive modeling (lecture 4 will illustrate this)

Thank you for your attention!