A case for systematic sound symbolism in pragmatics: Supporting information (All data)

Contents

Introduction

This is an analysis of an experiment into whether people can predict if an upcoming turn is a question or a statement, based on the previous turn type and the first phoneme of the target turn.

Participants listened to a series of audio samples. Each audio sample was made up of a *context* by speaker 1 (Statement or Inititating turn) and a *response* by speaker 2. The response was either no audio, a single segment [w] or a single semgent other than [w].

Load libraries

```
library(lme4)
library(lattice)
library(gplots)
library(ggplot2)
library(sjPlot)
library(party)
library(Rmisc)
library(dplyr)
```

Function for converting from logit scale

```
logit2per = function(X){
  return(exp(X)/(1+exp(X)))
}
```

Load data

```
d = read.csv("../Data/Lab_and_Online_data_Processed.csv")
```

Each row in the data is a single response from a participant to a single sample. The key variables are:

- *partID*: identifies participants
- *contextSample*: The name of the audio sample used for the context.
- *responseSample*: The name of the audio sample used for the response.
- *responsePhoneme*: The first segment of the response.
- *responseType*: Whether the first segment of the response came from a question or statement.
- *answer*: The participant's response to "Is the next turn a question?"

Make *answer* a binary variable.

```
d$answer = d$answer=="Yes"
d$lastAnswer = d$lastAnswer=="Yes"
```

Relevel response phoneme and context.

```
d$responsePhoneme = relevel(d$responsePhoneme, 'other')
d$context = relevel(d$context, 'ST')
```

Center trial number, so that the intercept will reflect probabilities in the middle of the experiment.

```
d$trialNumber.center = d$trialNumber - 25
```

Data exclusion

We exclude participant 13 because they took much longer than other participants.

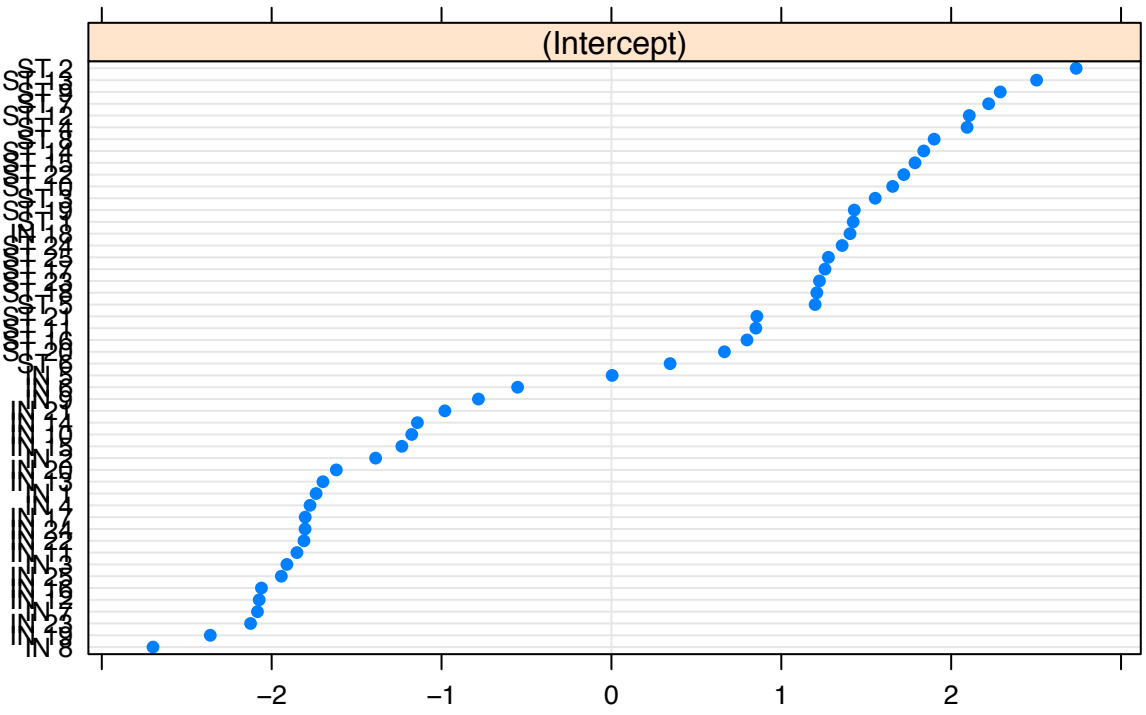
```
d = d[as.character(d$partID)!="13",]
```

Are there any samples that look like outliers? Make a basic model:

```
m3 = glmer(
  answer ~ 1 +
    (1  | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = glmerControl(optimizer="bobyqa", optCtrl = list(maxfun=2e4))
)
```

Then look at the random effects.

```
dotplot(ranef(m3))[[2]]
```



The sample "IN 18" is an outlier. However, models have convergence problems when leaving it out.

```
# Commented out - not run
#d = d[d$contextSample != 'IN 18',]
```

The data has 2400 observations:

```
# Number of observations per participant
table(d$partID)
```

```
##
##   1  10  11  12  13  14  15  16  17  18  19   2  20  21  22  23  24  25
##  50  50  50  50   0  50  50  50  50  50  50  50  50  50  50  50  50  50
##   3   4   5   6   7   8   9  01 010 011 012 013 014 015 016 017 018 019
##  50  50  50  50  50  50  50  50  50  50  50  50  50  50  50  50  50  50
##  02 020 021 022 023 024  03  04  05  06  07  08  09
##  50  50  50  50  50  50  50  50  50  50  50  50  50
```

```
table(d$context, d$responsePhoneme )
```

```
##
##       other none  wh
##   ST    488  123 589
##   IN    490  126 584
```

Exclude missing data

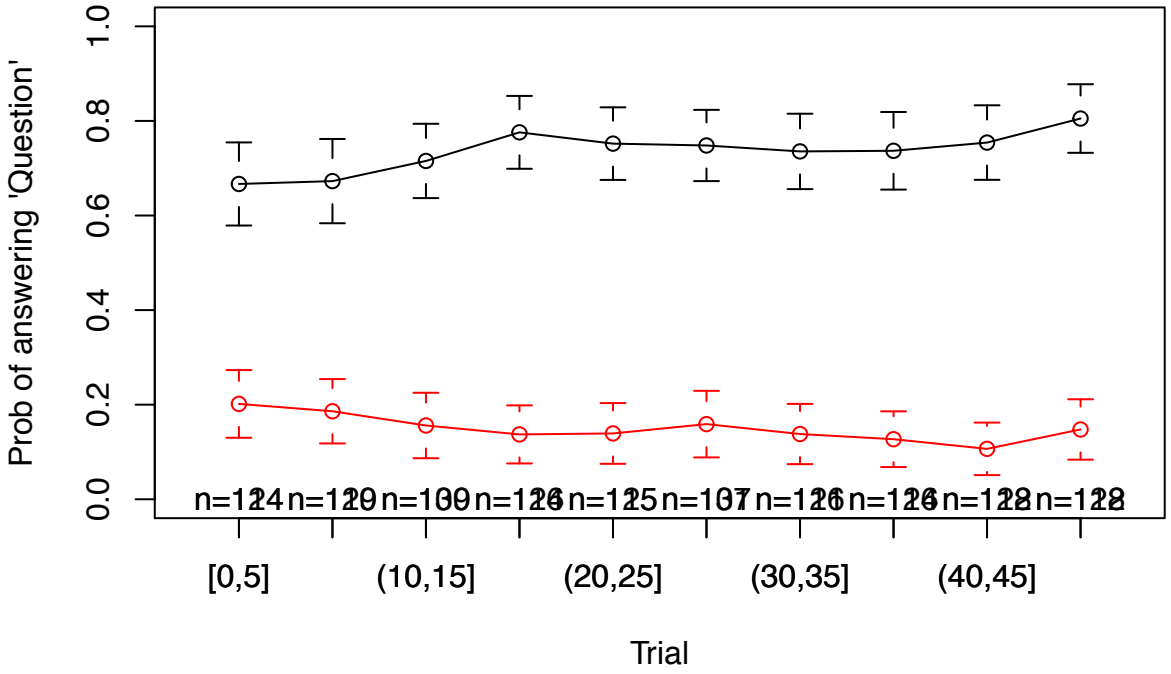
```
d = d[complete.cases(d[,c(
  "answer","trialNumber.center",
    "context","responsePhoneme",
  "context.sex","response.sex"
)]),]
```

Effects of block and trial

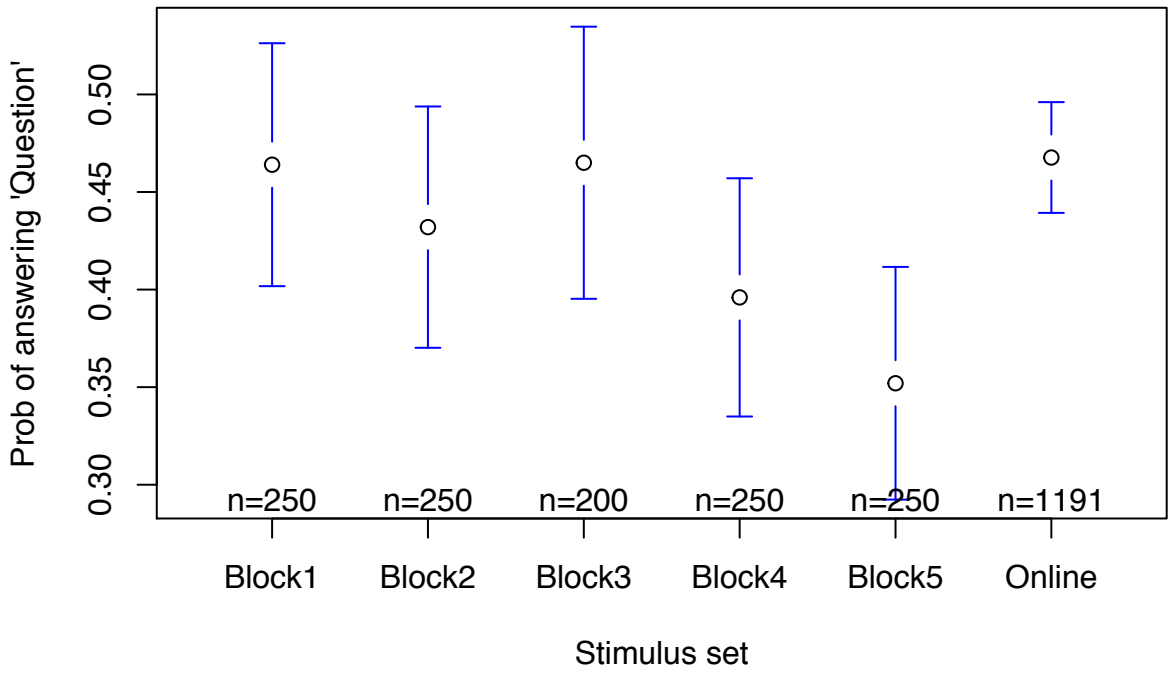
```
plotmeans(answer ~ cut(trialNumber,seq(0,50,length.out = 11), include.lowest = T),
          ylab = "Prob of answering 'Question'",
          xlab = 'Trial',
          data = d[d$context=="ST",],ylim=c(0,1),
          col = 1, barcol = 1)
plotmeans(answer ~ cut(trialNumber,seq(0,50,length.out = 11), include.lowest = T),
          ylab = "Prob of answering 'Question'",
          xlab = 'Trial',
          data = d[d$context=="IN",],ylim=c(0,1),
          col = 2, barcol = 2, add=T)
```

```
## Warning in axis(1, at = 1:length(means), labels = legends, ...): "add" is
## not a graphical parameter
```

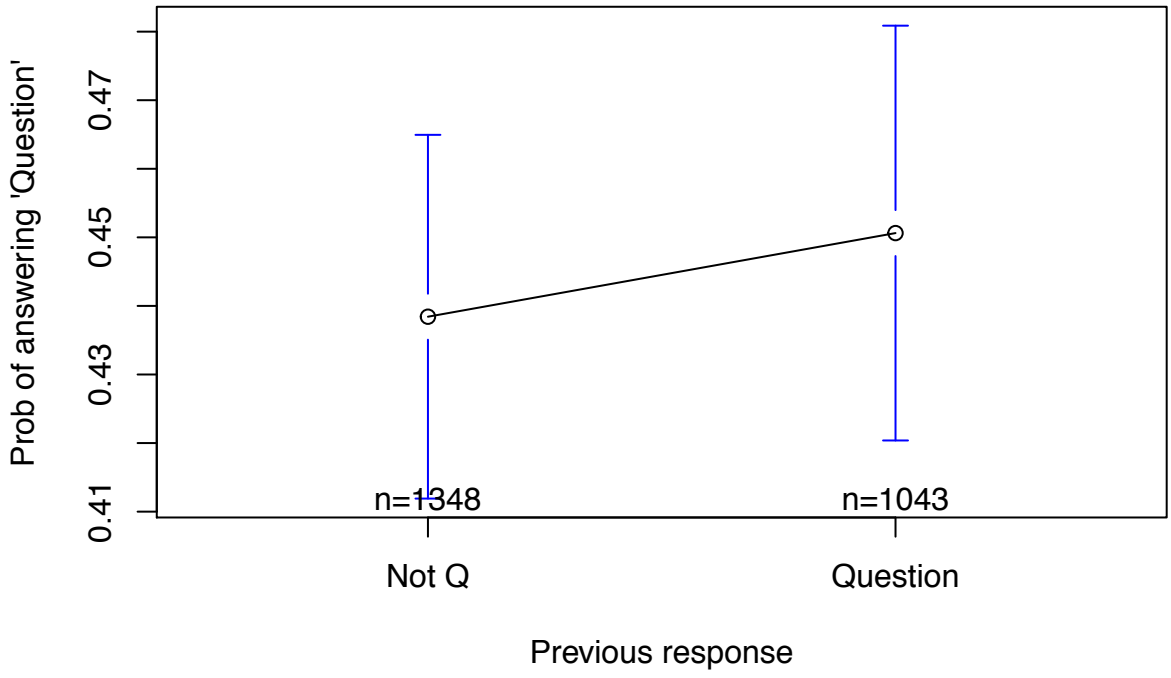
```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "add" is not a
## graphical parameter
```




```
plotmeans(d$answer ~ d$blockName,
          ylab = "Prob of answering 'Question'",
          xlab = 'Stimulus set', connect = F)
```




```
plotmeans(answer ~ lastAnswer,
          ylab = "Prob of answering 'Question'",
          xlab = "Previous response",
          legends = c("Not Q", "Question"),
          data = d)
```



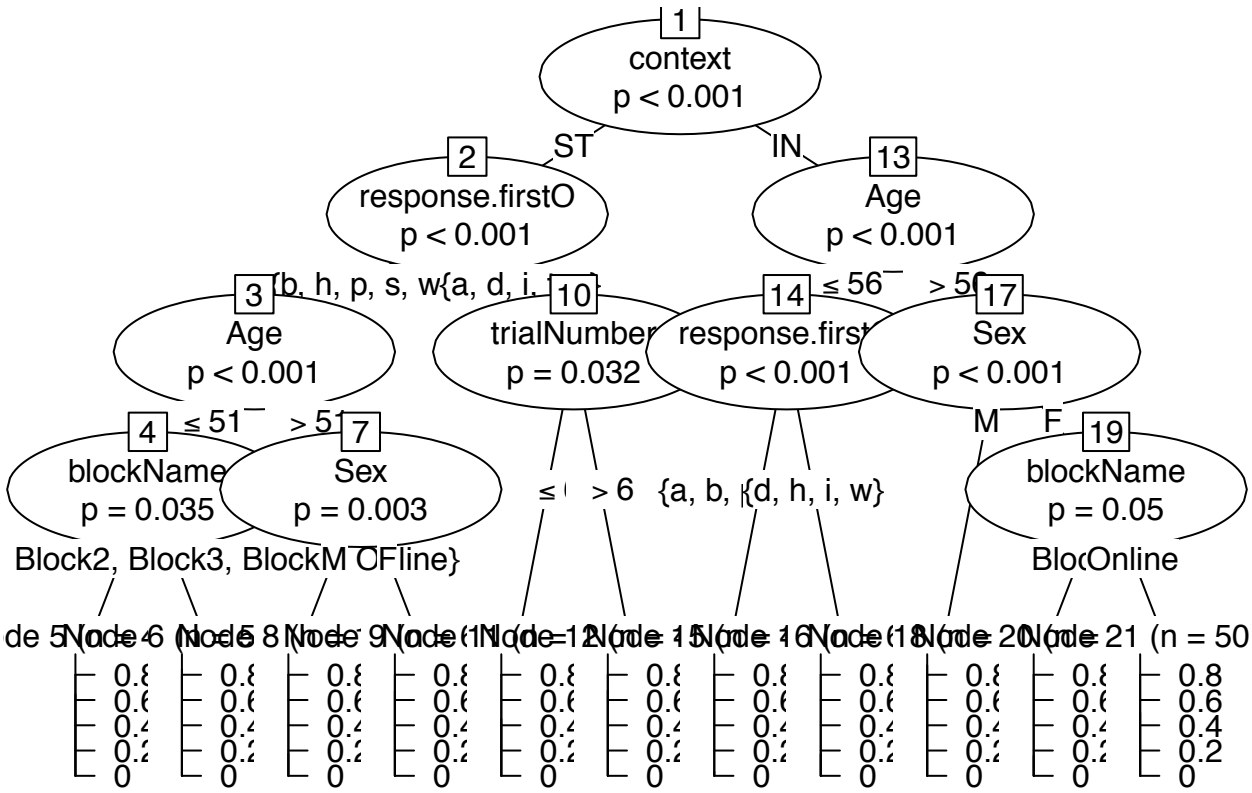
Decision tree

In order to get an idea of the structure of the data, we make a binary decision tree based on the data. We try to predict the participant's response by context, the first phoneme of the response, the type of turn the response was taken from and the participants' age and sex.

```
cx = ctree(answer ~
             context +responsePhoneme + responseType +
             Age + Sex +
             response.sex + context.sex +
             response.first0 + trialNumber + lastAnswer +
             blockName,
           data = d,
           controls = ctree_control(mincriterion = 0.95))
```

Plot the decision tree:

```
plot(cx, terminal_panel=node_barplot)
```



Context is the most important factor, followed by first phoneme of the response.

Mixed effects models

Make a series of mixed effects models. We can fix this using the "bobyqa" optimiser for both phases of the convergence and letting the algorithm run longer:

```
gcontrol = glmerControl(optimizer="bobyqa",optCtrl = list(maxfun=2e4))
```

(Note that several convergence algorithms were tested, and the three best fitting solutions had essentially no differences in fixed effect estimates)

Random effects structure

We have a good idea of what the random effects structure should be, but first we check whether there are significant differences by participant etc.

```
mA0 = glmer(
  answer ~ 1 +
    (1 | ESource /partID),
  data = d,
  family = binomial,
  control = gcontrol
)


mA0b = glmer(
  answer ~ 1 +
    (1 | blockName/partID) ,
  data = d,
  family = binomial,
  control = gcontrol
)
anova(mA0,mA0b)
```

```
## Data: d
## Models:
## mA0: answer ~ 1 + (1 | ESource/partID)
## mA0b: answer ~ 1 + (1 | blockName/partID)
##      Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## mA0   3 3288.8 3306.2 -1641.4   3282.8
## mA0b  3 3286.8 3304.2 -1640.4   3280.8 2.005      0  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is no significant improvement in the model when taking stimulus set into account. Because it complicates the analysis, we'll leave it out.

```
mA1 = glmer(
  answer ~ 1 +
    (1 | ESource /partID) +
    (1 | contextSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
mA2 = glmer(
  answer ~ 1 +
    (1 | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)

mA3 = glmer(
  answer ~ 1 +
    (1 + context| ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)

mA4 = glmer(
  answer ~ 1 +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 2 negative
## eigenvalues
```

```
anova(mA0, mA1, mA2, mA3, mA4)
```

```
## Data: d
## Models:
## mA0: answer ~ 1 + (1 | ESource/partID)
## mA1: answer ~ 1 + (1 | ESource/partID) + (1 | contextSample)
## mA2: answer ~ 1 + (1 | ESource/partID) + (1 | contextSample) + (1 |
## mA2:     responseSample)
## mA3: answer ~ 1 + (1 + context | ESource/partID) + (1 | contextSample) +
## mA3:     (1 | responseSample)
## mA4: answer ~ 1 + (1 + context | ESource/partID) + (0 + responsePhoneme |
## mA4:     ESource/partID) + (1 | contextSample) + (1 | responseSample)
##     Df    AIC    BIC   logLik deviance   Chisq Chi Df Pr(>Chisq)
## mA0  3 3288.8 3306.2 -1641.41   3282.8
## mA1  4 2356.8 2379.9 -1174.40   2348.8 934.028      1  < 2.2e-16 ***
## mA2  5 2326.4 2355.3 -1158.20   2316.4  32.393      1  1.259e-08 ***
## mA3  9 1948.8 2000.8  -965.39   1930.8 385.611      4  < 2.2e-16 ***
```

```
## mA4 21 1957.3 2078.7  -957.66   1915.3  15.472     12     0.2166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All proposed random effects significantly improve the fit of the model, except for the random slope for responsePhoneme by participant. This variable nearly doubles the number of model parameters, so we will leave it out.

Fixed effects

We are most interested in the effects of context and response type, but we need to check some other possible confounding variables.

Trial

```
m0 = glmer(
  answer ~ 1  +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 2 negative
## eigenvalues
```

```
trial = glmer(
  answer ~ 1 + trialNumber.center +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 2 negative
## eigenvalues
```

```
trialQ = glmer(
  answer ~ 1 + trialNumber.center + I(trialNumber.center^2) +
    (1 + context | ESource /partID) +
```

```
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.135616 (tol =
## 0.001, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unide
##  - Rescale variables?;Model is nearly unidentifiable: large eigenvalue ratio
##  - Rescale variables?
```

```
anova(m0,trial, trialQ)
```

```
## Data: d
## Models:
## m0: answer ~ 1 + (1 + context | ESource/partID) + (0 + responsePhoneme |
## m0:     ESource/partID) + (1 | contextSample) + (1 | responseSample)
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## trialQ: answer ~ 1 + trialNumber.center + I(trialNumber.center^2) + (1 +
## trialQ:     context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## trialQ:     (1 | contextSample) + (1 | responseSample)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0     21 1957.3 2078.7 -957.66   1915.3
## trial  22 1958.5 2085.6 -957.24   1914.5 0.8412      1    0.35906
## trialQ 23 1957.5 2090.4 -955.76   1911.5 2.9486      1    0.08595 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A significant effect of trial, but no significant quadratic term.

Previous answer

```
prevAns = glmer(
  answer ~ 1 + trialNumber.center + lastAnswer +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
anova(trial,prevAns)
```

```
## Data: d
## Models:
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## prevAns: answer ~ 1 + trialNumber.center + lastAnswer + (1 + context |
## prevAns:     ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## prevAns:     (1 | contextSample) + (1 | responseSample)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## trial   22 1958.5 2085.6 -957.24   1914.5
## prevAns 23 1960.1 2093.0 -957.04   1914.1 0.4032      1     0.5254
```

No significant effect of previous answer.

Sex of speakers in samples

```
contS = glmer(
  answer ~ 1 + trialNumber.center +
    context.sex +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 2 negative
## eigenvalues
```

```
respS = glmer(
  answer ~ 1 + trialNumber.center +
    context.sex + response.sex +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
```

```
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
contXrespS = glmer(
  answer ~ 1 + trialNumber.center +
    context.sex * response.sex +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
anova(trial,contS, respS, contXrespS)
```

```
## Data: d
## Models:
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## contS: answer ~ 1 + trialNumber.center + context.sex + (1 + context |
## contS:     ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## contS:     (1 | contextSample) + (1 | responseSample)
## respS: answer ~ 1 + trialNumber.center + context.sex + response.sex +
## respS:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## respS:     (1 | contextSample) + (1 | responseSample)
## contXrespS: answer ~ 1 + trialNumber.center + context.sex * response.sex +
## contXrespS:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## contXrespS:     (1 | contextSample) + (1 | responseSample)
##            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## trial      22 1958.5 2085.6 -957.24   1914.5
## contS      23 1957.8 2090.7 -955.91   1911.8 2.6638      1     0.1027
## respS      24 1959.6 2098.3 -955.79   1911.6 0.2298      1     0.6317
## contXrespS 25 1961.6 2106.1 -955.81   1911.6 0.0000      1     1.0000
```

No significant effects of the sex of the speakers in the samples.

Sex of participants

```
sex = glmer(
  answer ~ 1 + trialNumber.center + Sex +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
anova(trial,sex)
```

```
## Data: d
## Models:
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## sex: answer ~ 1 + trialNumber.center + Sex + (1 + context | ESource/partID) +
## sex:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## sex:     (1 | responseSample)
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## trial 22 1958.5 2085.6 -957.24   1914.5
## sex   24 1959.5 2098.2 -955.77   1911.5 2.9322      2     0.2308
```

No significant effect of the sex of the participant.

Effects of Context and Response

The only significant confounding variable is trial.

```
context = glmer(
  answer ~ 1 + trialNumber.center +
    context +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
```

```
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.102715 (tol =
## 0.001, component 1)
```

```
rPhon = glmer(
  answer ~ 1 + trialNumber.center +
    context + responsePhoneme +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.127637 (tol =
## 0.001, component 1)
```

```
conXrPh = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.152948 (tol =
## 0.001, component 1)
```

```
anova(trial, context,rPhon, conXrPh)
```

```
## Data: d
## Models:
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## context: answer ~ 1 + trialNumber.center + context + (1 + context | ESource/partID) +
## context:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## context:     (1 | responseSample)
## rPhon: answer ~ 1 + trialNumber.center + context + responsePhoneme +
## rPhon:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
```

```
## rPhon:     (1 | contextSample) + (1 | responseSample)
## conXrPh: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## conXrPh:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## conXrPh:     (1 | contextSample) + (1 | responseSample)
##         Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## trial   22 1958.5 2085.6 -957.24   1914.5
## context 23 1949.4 2082.3 -951.69   1903.4 11.0907      1  0.0008676 ***
## rPhon   25 1948.7 2093.2 -949.34   1898.7  4.6988      2  0.0954277 .
## conXrPh 27 1951.0 2107.0 -948.50   1897.0  1.6917      2  0.4291956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction between Sex and responses

```
Sex = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    Sex +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.177809 (tol =
## 0.001, component 1)
```

```
SexXresp = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    Sex*responsePhoneme +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
SexXcon = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    Sex*responsePhoneme +
    Sex:context +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.15954 (tol =
## 0.001, component 1)
```

```
SxXcoXre = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    Sex*responsePhoneme*context +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample) ,
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## fixed-effect model matrix is rank deficient so dropping 2 columns / coefficients
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.122017 (tol =
## 0.001, component 1)
```

```
anova(conXrPh, Sex, SexXresp, SexXcon, SxXcoXre)
```

```
## Data: d
## Models:
## conXrPh: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## conXrPh:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## conXrPh:     (1 | contextSample) + (1 | responseSample)
## Sex: answer ~ 1 + trialNumber.center + context * responsePhoneme +
```

```
## Sex:     Sex + (1 + context | ESource/partID) + (0 + responsePhoneme |
## Sex:     ESource/partID) + (1 | contextSample) + (1 | responseSample)
## SexXresp: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## SexXresp:     Sex * responsePhoneme + (1 + context | ESource/partID) +
## SexXresp:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## SexXresp:     (1 | responseSample)
## SexXcon: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## SexXcon:     Sex * responsePhoneme + Sex:context + (1 + context | ESource/partID) +
## SexXcon:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## SexXcon:     (1 | responseSample)
## SxXcoXre: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## SxXcoXre:     Sex * responsePhoneme * context + (1 + context | ESource/partID) +
## SxXcoXre:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## SxXcoXre:     (1 | responseSample)
##          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## conXrPh  27 1951.0 2107.0 -948.50   1897.0
## Sex      29 1952.4 2120.0 -947.20   1894.4 2.5858      2     0.2745
## SexXresp 32 1957.4 2142.4 -946.72   1893.4 0.9650      3     0.8097
## SexXcon  34 1960.4 2156.9 -946.22   1892.4 1.0091      2     0.6038
## SxXcoXre 37 1965.4 2179.2 -945.68   1891.4 1.0725      3     0.7837
```

No effect by sex of participant.

Interaction with trial

```
trialXCon = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    trialNumber.center:context +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.198699 (tol =
## 0.001, component 1)
```

```
trialXph = glmer(
  answer ~ 1 + trialNumber.center +
    context * responsePhoneme +
    trialNumber.center:context +
    trialNumber.center:responsePhoneme +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
```

```
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.14431 (tol =
## 0.001, component 1)
```

```
trXcoXph = glmer(
  answer ~ 1 + trialNumber.center *
    context * responsePhoneme +
    (1 + context | ESource /partID) +
    (0 + responsePhoneme | ESource /partID) +
    (1 | contextSample) +
    (1 | responseSample),
  data = d,
  family = binomial,
  control = gcontrol
)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.227 (tol = 0.001,
## component 1)
```

```
anova(conXrPh, trialXCon, trialXph, trXcoXph)
```

```
## Data: d
## Models:
## conXrPh: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## conXrPh:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## conXrPh:     (1 | contextSample) + (1 | responseSample)
## trialXCon: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## trialXCon:     trialNumber.center:context + (1 + context | ESource/partID) +
## trialXCon:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trialXCon:     (1 | responseSample)
## trialXph: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## trialXph:     trialNumber.center:context + trialNumber.center:responsePhoneme +
## trialXph:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## trialXph:     (1 | contextSample) + (1 | responseSample)
## trXcoXph: answer ~ 1 + trialNumber.center * context * responsePhoneme +
## trXcoXph:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## trXcoXph:     (1 | contextSample) + (1 | responseSample)
##           Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## conXrPh   27 1951.0 2107.0 -948.50   1897.0
## trialXCon 28 1939.8 2101.7 -941.92   1883.8 13.1538      1  0.0002869 ***
## trialXph  30 1942.7 2116.1 -941.35   1882.7  1.1398      2  0.5655917
## trXcoXph  32 1945.2 2130.1 -940.59   1881.2  1.5167      2  0.4684470
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results

Model comparison

```
anova(trial, context,rPhon, conXrPh)
```

```
## Data: d
## Models:
## trial: answer ~ 1 + trialNumber.center + (1 + context | ESource/partID) +
## trial:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## trial:     (1 | responseSample)
## context: answer ~ 1 + trialNumber.center + context + (1 + context | ESource/partID) +
## context:     (0 + responsePhoneme | ESource/partID) + (1 | contextSample) +
## context:     (1 | responseSample)
## rPhon: answer ~ 1 + trialNumber.center + context + responsePhoneme +
## rPhon:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## rPhon:     (1 | contextSample) + (1 | responseSample)
## conXrPh: answer ~ 1 + trialNumber.center + context * responsePhoneme +
## conXrPh:     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
## conXrPh:     (1 | contextSample) + (1 | responseSample)
##         Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## trial   22 1958.5 2085.6 -957.24   1914.5
## context 23 1949.4 2082.3 -951.69   1903.4 11.0907      1  0.0008676 ***
## rPhon   25 1948.7 2093.2 -949.34   1898.7  4.6988      2  0.0954277 .
## conXrPh 27 1951.0 2107.0 -948.50   1897.0  1.6917      2  0.4291956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fixed effects

Model estimates:

```
finalModel = conXrPh
save(finalModel, file="../results/FinalModel.Rdat")
summary(finalModel)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: answer ~ 1 + trialNumber.center + context * responsePhoneme +
##     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
##     (1 | contextSample) + (1 | responseSample)
##    Data: d
## Control: gcontrol
##
##      AIC      BIC   logLik deviance df.resid
##   1951.0   2107.0   -948.5   1897.0     2364
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -19.6248  -0.3088  -0.1310   0.4161   8.9024
##
## Random effects:
##  Groups           Name                 Variance  Std.Dev. Corr
##  responseSample   (Intercept)          1.880e-01 0.433566
##  contextSample    (Intercept)          9.647e-01 0.982212
##  partID.ESource   responsePhonemeother 8.936e-02 0.298934
##                   responsePhonemenone  6.169e-01 0.785454 0.94
##                   responsePhonemewh    1.332e-01 0.364961 0.55  0.80
##  partID.ESource.1 (Intercept)          1.597e+00 1.263695
##                   contextIN            6.330e+00 2.516035 -1.00
##  ESource          responsePhonemeother 4.611e-02 0.214732
##                   responsePhonemenone  1.057e-02 0.102815 1.00
##                   responsePhonemewh    1.347e-04 0.011605 0.99  0.99
##  ESource.1        (Intercept)          1.228e-06 0.001108
##                   contextIN            1.132e-05 0.003364 -1.00
## Number of obs: 2391, groups:
## responseSample, 51; contextSample, 50; partID:ESource, 48; ESource, 2
##
## Fixed effects:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    0.967890   0.344109   2.813 0.004912 **
## trialNumber.center             0.004183   0.004304   0.972 0.331174
## contextIN                     -3.808757   0.514774  -7.399 1.37e-13 ***
## responsePhonemenone           -0.287864   0.780094  -0.369 0.712119
## responsePhonemewh              0.914794   0.262074   3.491 0.000482 ***
## contextIN:responsePhonemenone -0.727367   0.589811  -1.233 0.217494
## contextIN:responsePhonemewh   -0.090541   0.283234  -0.320 0.749220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##                    (Intr) trlNm. cntxIN rspnsPhnmn rspnsPhnmw
```

```
## trlNmbr.cnt        -0.001
## contextIN          -0.677 -0.005
## rspnsPhnmnn        -0.113  0.011  0.031
## rspnsPhnmwh        -0.478  0.003  0.091  0.157
## cntxtIN:rspnsPhnmn  0.068 -0.031 -0.152 -0.167     -0.071
## cntxtIN:rspnsPhnmw  0.142 -0.021 -0.325 -0.051     -0.413
##                    cntxtIN:rspnsPhnmn
## trlNmbr.cnt
## contextIN
## rspnsPhnmnn
## rspnsPhnmwh
## cntxtIN:rspnsPhnmn
## cntxtIN:rspnsPhnmw  0.256
## convergence code: 1
## Model failed to converge with max|grad| = 0.152948 (tol = 0.001, component 1)
```

Relevel the response phoneme to see other comparisons:

```
d2 = d
d2$responsePhoneme = relevel(d2$responsePhoneme,"wh")
fm2 = update(finalModel, data=d2)
```

```
## Warning in optwrap(optimizer, devfun, start, rho$lower, control =
## control, : convergence code 1 from bobyqa: bobyqa -- maximum number of
## function evaluations exceeded
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.164989 (tol =
## 0.001, component 1)
```

```
summary(fm2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: answer ~ 1 + trialNumber.center + context * responsePhoneme +
##     (1 + context | ESource/partID) + (0 + responsePhoneme | ESource/partID) +
##     (1 | contextSample) + (1 | responseSample)
##    Data: d2
## Control: gcontrol
##
##      AIC      BIC   logLik deviance df.resid
##   1951.0   2107.0   -948.5   1897.0     2364
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -19.6855  -0.3093  -0.1309   0.4168   8.9243
##
## Random effects:
##  Groups           Name                 Variance  Std.Dev.  Corr
##  responseSample   (Intercept)          1.866e-01 0.4319450
##  contextSample    (Intercept)          9.652e-01 0.9824284
##  partID.ESource   responsePhonemewh    1.313e-01 0.3623206
##                   responsePhonemeother 9.021e-02 0.3003559 0.56
##                   responsePhonemenone  6.268e-01 0.7916756 0.80  0.94
##  partID.ESource.1 (Intercept)          1.590e+00 1.2607934
```

```
##                   contextIN            6.312e+00 2.5123304 -1.00
##  ESource          responsePhonemewh    1.413e-04 0.0118856
##                   responsePhonemeother 4.705e-02 0.2168989 1.00
##                   responsePhonemenone  1.314e-02 0.1146227 0.99  0.99
##  ESource.1        (Intercept)          4.450e-07 0.0006671
##                   contextIN            6.497e-06 0.0025489 -1.00
## Number of obs: 2391, groups:
## responseSample, 51; contextSample, 50; partID:ESource, 48; ESource, 2
##
## Fixed effects:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     1.871022   0.317036   5.902 3.60e-09 ***
## trialNumber.center              0.004201   0.004303   0.976 0.328948
## contextIN                      -3.883274   0.499837  -7.769 7.91e-15 ***
## responsePhonemeother           -0.897820   0.262071  -3.426 0.000613 ***
## responsePhonemenone            -1.168504   0.797953  -1.464 0.143091
## contextIN:responsePhonemeother  0.071944   0.283188   0.254 0.799457
## contextIN:responsePhonemenone  -0.680547   0.588585  -1.156 0.247582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##                   (Intr) trlNm. cntxIN rspnsPhnmt rspnsPhnmn
## trlNmbr.cnt         0.001
## contextIN          -0.784 -0.017
## rspnsPhnmth        -0.303 -0.003  0.139
## rspnsPhnmnn        -0.092  0.011  0.048  0.187
## cntxtIN:rspnsPhnmt  0.187  0.020 -0.231 -0.414     -0.086
## cntxtIN:rspnsPhnmn  0.104 -0.021 -0.122 -0.126     -0.181
##                   cntxtIN:rspnsPhnmt
## trlNmbr.cnt
## contextIN
## rspnsPhnmth
## rspnsPhnmnn
## cntxtIN:rspnsPhnmt
## cntxtIN:rspnsPhnmn  0.224
## convergence code: 1
## Model failed to converge with max|grad| = 0.164989 (tol = 0.001, component 1)
```

Confidence intervals (through Wald method):

```
CI = confint(finalModel,parm="beta_", method="Wald")
cx = summary(finalModel)$coef
cx = cbind(cx[,1],CI,cx[,2:4])
cx2 = cx
for(i in 1:5){cx2[,i] = round(cx2[,i],3)}
cx2
```

```
##                                    2.5 % 97.5 % Std. Error z value
## (Intercept)                     0.968  0.293  1.642      0.344   2.813
## trialNumber.center              0.004 -0.004  0.013      0.004   0.972
## contextIN                      -3.809 -4.818 -2.800      0.515  -7.399
## responsePhonemenone            -0.288 -1.817  1.241      0.780  -0.369
## responsePhonemewh               0.915  0.401  1.428      0.262   3.491
## contextIN:responsePhonemenone -0.727 -1.883  0.429      0.590  -1.233
```

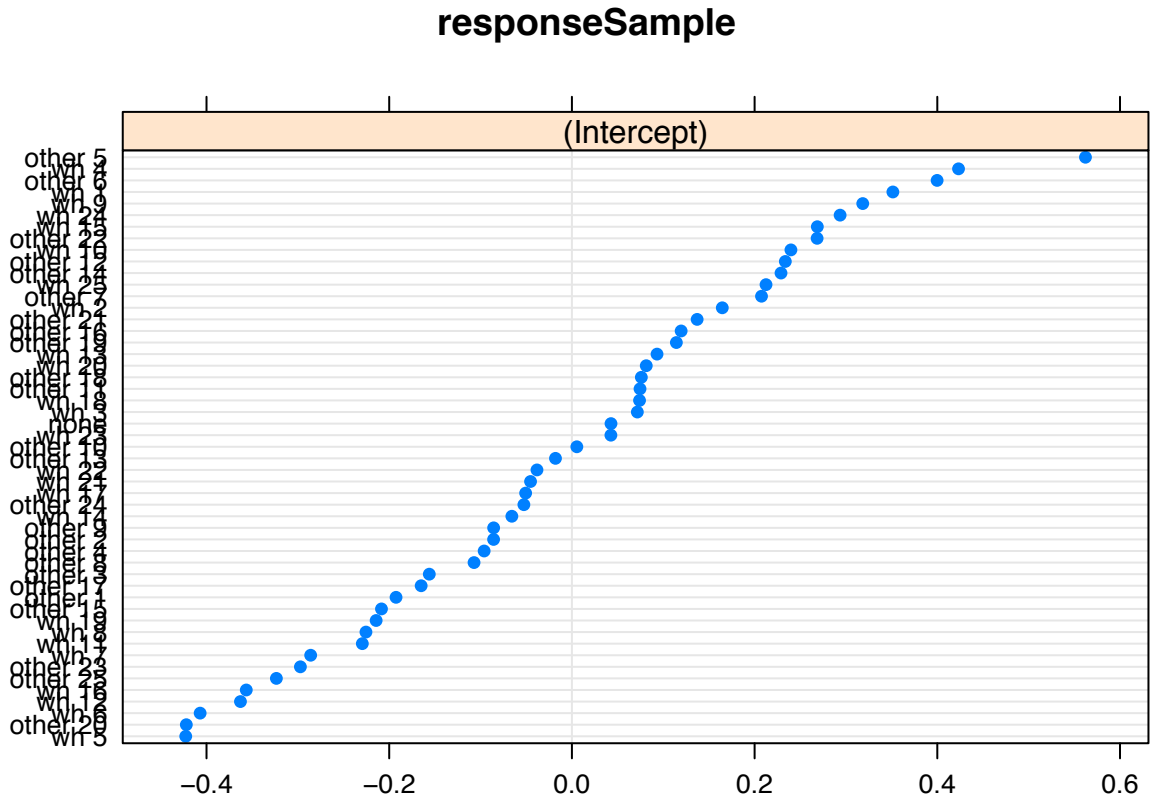
```
## contextIN:responsePhonemewh   -0.091 -0.646  0.465      0.283  -0.320
##                                   Pr(>|z|)
## (Intercept)                   4.912023e-03
## trialNumber.center            3.311742e-01
## contextIN                     1.373202e-13
## responsePhonemenone           7.121189e-01
## responsePhonemewh             4.819617e-04
## contextIN:responsePhonemenone 2.174937e-01
## contextIN:responsePhonemewh   7.492203e-01
```

```
write.csv(cx, "../results/FinalModelCoefficients.csv")
```

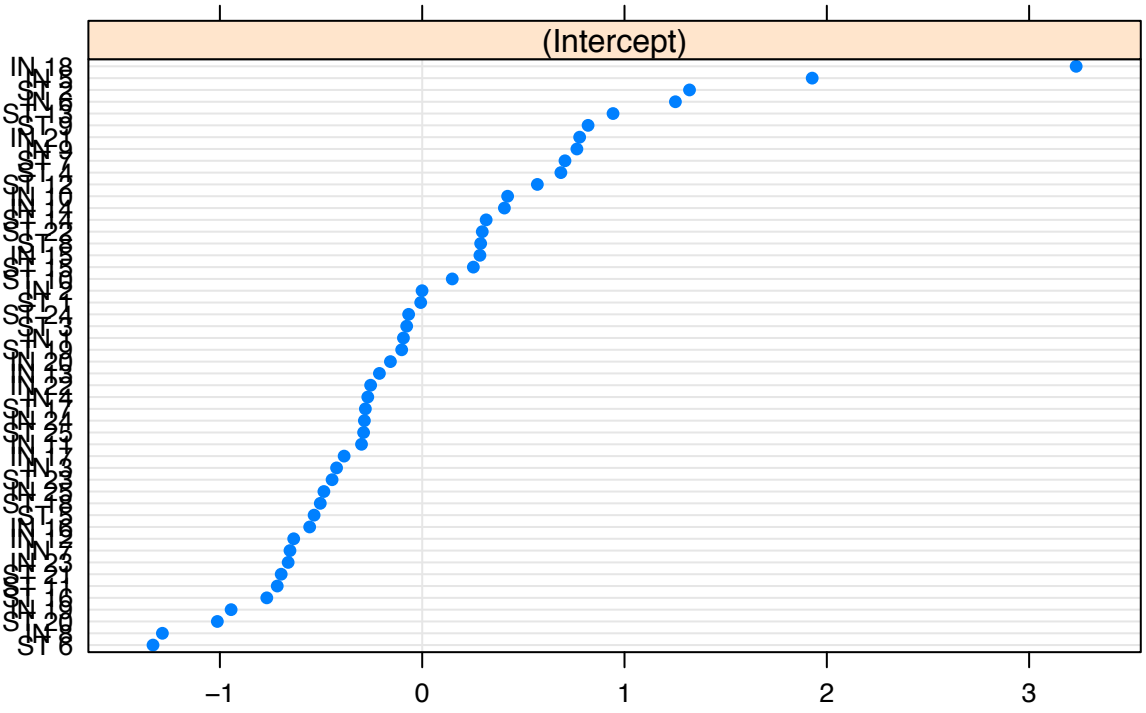
Random effects

```
dotplot(ranef(finalModel))
```

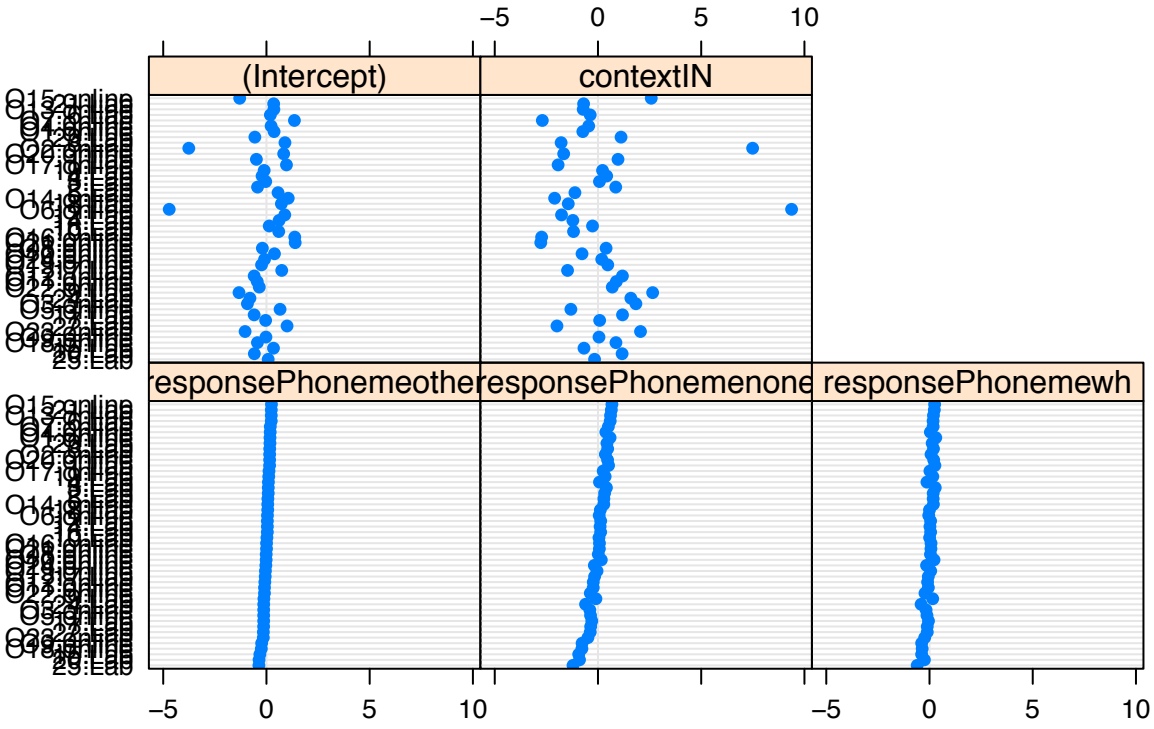
```
## $responseSample
```



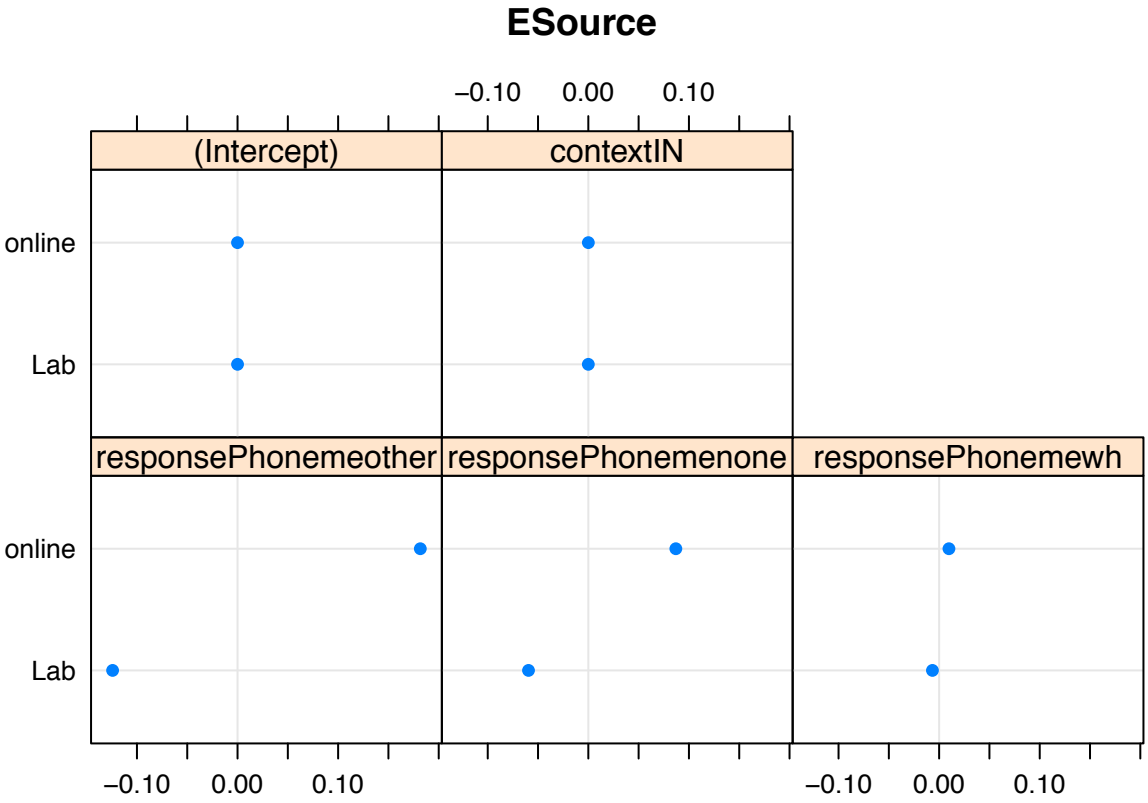

```
##
## $contextSample
```




```
##
## $`partID:ESource`
```



```
##
## $ESource
```


ESource
−0.10
0.00
0.10
(Intercept)
contextIN
online
Lab
responsePhonemeother
responsePhonemenone
responsePhonemewh
online
Lab
−0.10
0.00
0.10
−0.10
0.00
0.10

Summary

Here is a summary of the main results:

There was a significant main effect of context (log likelihood difference = 6 , df = 2 , Chi Squared = 11.93 , p = 0.0026).

There was a marginal main effect of phoneme (log likelihood difference = 2.3 , df = 2 , Chi Squared = 4.7 , p = 0.095).

There was no significant interaction between context and phoneme (log likelihood difference = 0.85 , df = 2 , Chi Squared = 1.69 , p = 0.43).

Work out model esimates for probabilities in each condition:

```
# prob of responding 'yes' when:
# Context = ST, other response
logit2per(fixef(finalModel)[1])[[1]]
```

```
## [1] 0.7246988
```

```
# Context = ST, no response
logit2per(fixef(finalModel)[1] + fixef(finalModel)["responsePhonemenone"]  )[[1]]
```

```
## [1] 0.6637446
```

```
# Context = ST, wh
logit2per(fixef(finalModel)[1] + fixef(finalModel)["responsePhonemewh"]  )[[1]]
```

```
## [1] 0.8679191
```

```
# Context = IN, other response
logit2per(fixef(finalModel)[1] + fixef(finalModel)["contextIN"])[[1]]
```

```
## [1] 0.05515538
```

```
# Context = IN, no response
logit2per(fixef(finalModel)[1] +
            fixef(finalModel)["contextIN"] +
            fixef(finalModel)["responsePhonemenone"])[[1]]
```

```
## [1] 0.0419376
```

```
# Context = IN, wh
logit2per(fixef(finalModel)[1] +
            fixef(finalModel)["contextIN"] +
            fixef(finalModel)["responsePhonemewh"])[[1]]
```

```
## [1] 0.1271859
```

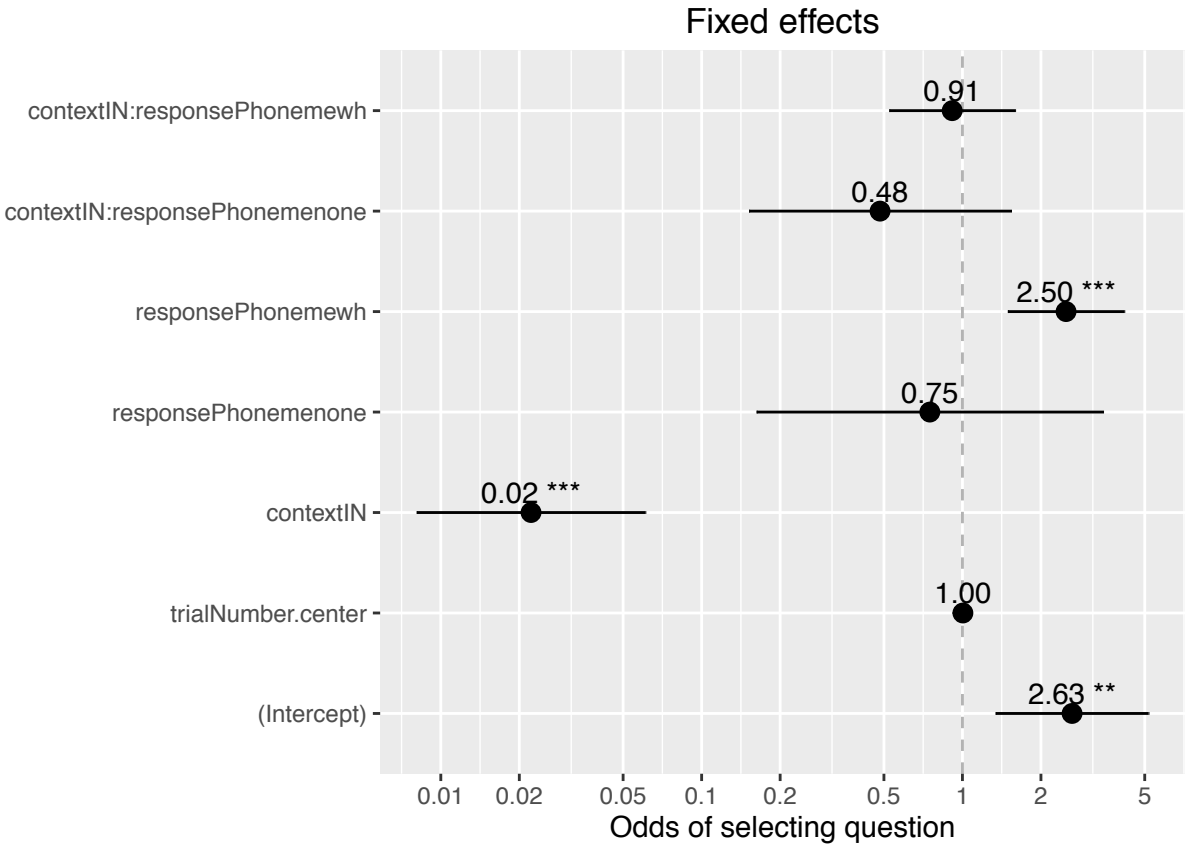
Plots

Fixed effects estimates:

```
feLabels = matrix(c(
  "(Intercept)"              ,"Intercept"      , NA,
  "trialNumber.center", "Trial",NA,
  "contextST", "Context = Statement", "context",
  "contextIN", "Context = Initiating", "context",
  "responsePhonemenone", "no response", 'rPhon',
  "responsePhonemewh", "wh response", 'rPhon',
  "contextIN:responsePhonemenone", "Context = In: no response", "conXrPh",
  "contextIN:responsePhonemewh", "Context = In: wh response", "conXrPh"
), ncol=3, byrow = T)

feLabels2 = as.vector(feLabels[match(names(fixef(finalModel)),feLabels[,1]),2])
```

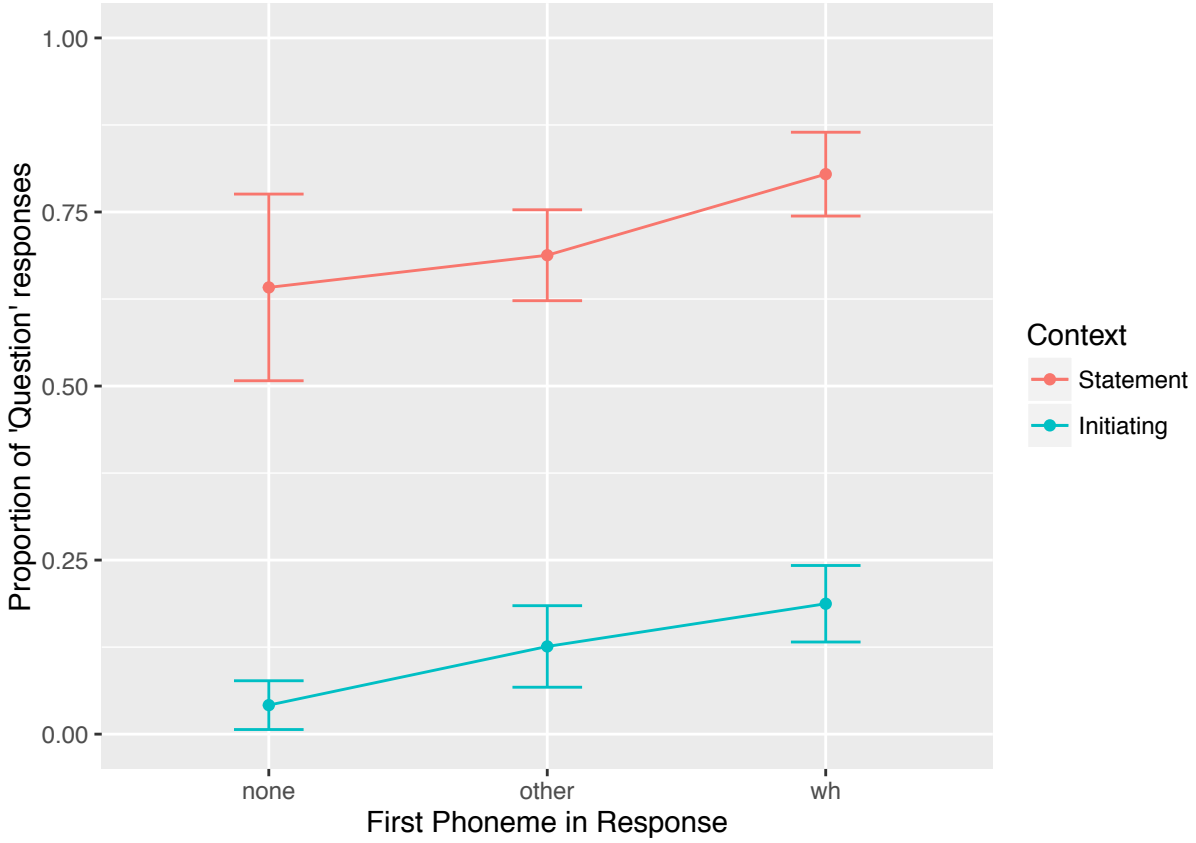
```
sjp.glmer(finalModel,'fe',
          show.intercept = T,
          geom.colors = c(1,1),
          axis.title = "Odds of selecting question",
          y.offset = 0.2
)
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```



Raw data plots

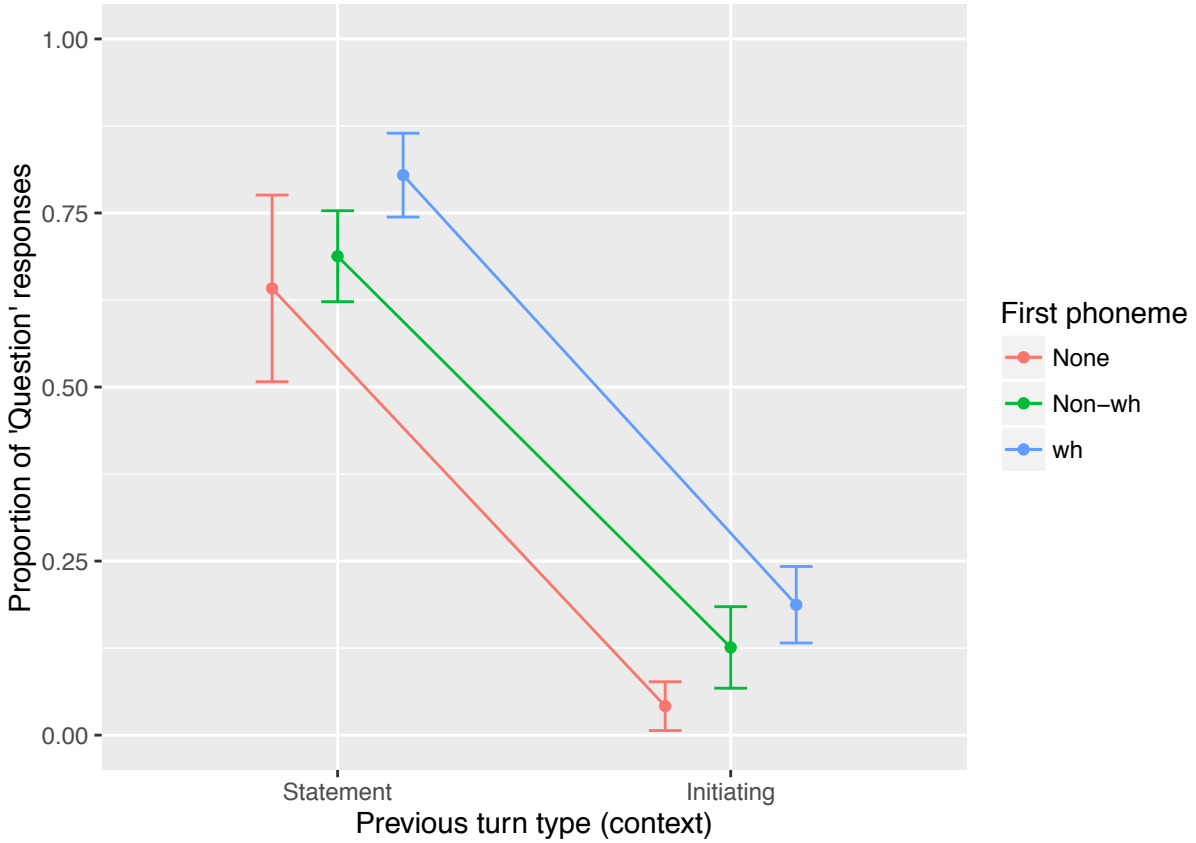
```
d$responsePhoneme = relevel(d$responsePhoneme,'none')

sumStats = group_by(d, partID ,context,responsePhoneme ) %>%
              summarise(mean =mean(answer))

sumStats2 = summarySE(sumStats, measurevar="mean", groupvars=c("context","responsePhoneme"))
sumStats2$upper = sumStats2$mean + sumStats2$ci
sumStats2$lower = sumStats2$mean - sumStats2$ci

dodge <- position_dodge(width=0.5)

main.plot <- ggplot(sumStats2,
    aes(x = responsePhoneme, y = mean, colour=context)) +
  geom_point() + geom_line(aes(group=context)) +
  geom_errorbar(aes(ymax=mean+ci, ymin=mean-ci), width=0.25) +
  xlab("First Phoneme in Response") +
  ylab("Proportion of 'Question' responses") +
  coord_cartesian(ylim=c(0,1)) +
  scale_color_discrete(breaks=c("ST","IN"),
                       labels=c("Statement","Initiating"),
                       name="Context")

main.plot
```



```
pdf("../results/graphs/PropQResponses_by_firstPhoneme_withPartCI.pdf",
    width = 4, height=3)
main.plot
dev.off()
```

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

```
main.plot2 <- ggplot(sumStats2,
    aes(x = context, y = mean, colour=responsePhoneme)) +
  geom_point(position=dodge) + geom_line(aes(group=responsePhoneme), position=dodge) +
  geom_errorbar(aes(ymax=mean+ci, ymin=mean-ci), width=0.25, position=dodge) +
  xlab("Previous turn type (context)") +
  ylab("Proportion of 'Question' responses") +
  coord_cartesian(ylim=c(0,1)) +
  scale_color_discrete(breaks=c("none","other",'wh'),
                       labels=c("None","Non-wh","wh"),
                       name="First phoneme") +
  scale_x_discrete(breaks=c("ST", "IN"),
                   labels=c("Statement", "Initiating"))

main.plot2
```




```
pdf("../results/graphs/PropQResponses_by_context_withPartCI.pdf",
    width = 4, height=3)
main.plot2
dev.off()
```

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