Improving memory for and production of singular they pronouns: Experiment 1

Bethany Gardner

03/24/2022

Contents

Load Data

Read data, preprocessed from Qualtrics output. See data/exp1_data_readme for more details.

```
d <- read.csv("../data/exp1_data.csv", stringsAsFactors=TRUE)
str(d)
```

```
## 'data.frame':    5202 obs. of  17 variables:
##  $ SubjID     : Factor w/ 102 levels "R_0qPfWjp8o4W3Z61",..: 84 84 84 84 84 84 84 84 84 84 ...
##  $ SubjAge    : int  18 18 18 18 18 18 18 18 18 18 ...
##  $ SubjGender : Factor w/ 3 levels "female","male",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ SubjEnglish: Factor w/ 3 levels "Fully competent in speaking listening reading and writing but no
##  $ List       : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ Task       : Factor w/ 3 levels "introduction",..: 2 2 2 3 2 2 2 3 2 2 ...
##  $ Name       : Factor w/ 12 levels "Amanda","Andrew",..: 3 3 3 3 8 8 8 8 4 4 ...
##  $ Pronoun    : Factor w/ 3 levels "he/him","she/her",..: 1 1 1 1 1 1 1 1 3 3 ...
##  $ Pet        : Factor w/ 3 levels "cat","dog","fish": 1 1 1 1 2 2 2 2 3 3 ...
```

```
##  $ Job        : Factor w/ 12 levels "accountant","doctor",..: 12 12 12 12 4 4 4 4 9 9 ...
##  $ M_Type     : Factor w/ 4 levels "","job","pet",..: 2 3 4 1 2 3 4 1 2 3 ...
##  $ M_Response : Factor w/ 19 levels "","accountant",..: 19 3 9 1 8 5 18 1 14 7 ...
##  $ M_Acc      : int  0 1 1 NA 1 1 0 NA 1 1 ...
##  $ P_Response : Factor w/ 793 levels "","amanda fixed her computer",..: 1 1 1 169 1 1 1 235 1 1 ...
##  $ P_Pronoun  : Factor w/ 5 levels "","he/him","none",..: 1 1 1 2 1 1 1 2 1 1 ...
##  $ P_Acc      : int  NA NA NA 1 NA NA NA 1 NA NA ...
##  $ I_Response : Factor w/ 299 levels "","  I would ask Jessica about food spots ",..: 1 1 1 1 1 1 1 1
```

Set up contrast coding. The first contrast compares they to he+she. The second contrast compares he to she.

```
contrasts(d$Pronoun)=cbind("they vs he+she"=c(.33,.33,-.66),
                           "he vs she"=c(-.5,.5, 0))
contrasts(d$Pronoun)
```

```
##           they vs he+she he vs she
## he/him              0.33      -0.5
## she/her             0.33       0.5
## they/them          -0.66       0.0
```

Split data by task, and only keep pronoun questions (not the job or pet questions) in memory dataframe.

```
m <- d %>% filter(M_Type=="pronoun")
p <- d %>% filter(Task=="production")
```

Combine memory and production trials to make one row for each character.

```
#Get pronoun memory and production observations. Filter out memory for job and pet questions, and intro
mp <- d %>% filter(Task != "introduction" &
                   M_Type != "job" &
                   M_Type != "pet")

#Just take columns used in model
m_temp <- mp %>% select(M_Acc, Pronoun, Name, SubjID) %>%
  filter(!is.na(M_Acc)) #Take out empty rows that were other question types

#Get production accuracy column
p_temp <- mp %>% select(P_Acc)  %>%
    filter(!is.na(P_Acc)) #Take out empty rows that were other question types

#Combine
mp <- cbind(m_temp, p_temp)
str(mp)
```

```
## 'data.frame':    1224 obs. of  5 variables:
##  $ M_Acc  : int  1 0 1 1 1 1 1 1 1 1 ...
##  $ Pronoun: Factor w/ 3 levels "he/him","she/her",..: 1 1 3 3 2 2 1 2 2 1 ...
##   ..- attr(*, "contrasts")= num [1:3, 1:2] 0.33 0.33 -0.66 -0.5 0.5 0
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:3] "he/him" "she/her" "they/them"
##   .. .. ..$ : chr [1:2] "they vs he+she" "he vs she"
##  $ Name   : Factor w/ 12 levels "Amanda","Andrew",..: 3 8 4 9 10 12 2 5 1 11 ...
##  $ SubjID : Factor w/ 102 levels "R_0qPfWjp8o4W3Z61",..: 84 84 84 84 84 84 84 84 84 84 ...
##  $ P_Acc  : int  1 1 0 0 1 1 1 1 1 1 ...
```

Memory

Descriptive Stats

Mean accuracy for all three memory question types.

```
prop.table(table(d$M_Type, d$M_Acc), margin=1)
```

```
##
##                   0         1
##
##   job     0.7867647 0.2132353
##   pet     0.5890523 0.4109477
##   pronoun 0.3423203 0.6576797
```

Mean accuracy, split by pronoun type.

```
prop.table(table(m$Pronoun, m$M_Acc), margin=1)
```

```
##
##                     0         1
##   he/him    0.2377451 0.7622549
##   she/her   0.2254902 0.7745098
##   they/them 0.5637255 0.4362745
```

94% of participants selected they/them at least once.

```
they_m <- d %>% filter(M_Response=="they/them") %>%
  summarize(n=n_distinct(SubjID))

they_m/(n_distinct(d$SubjID))
```

```
##           n
## 1 0.9411765
```

Model

Start with model that has random intercepts and slopes for participant and item. Specifying the direction as "order" in buildmer will identify the maximal model that will converge. However, it doesn't continue to backward stepwise elimination. This results in a model with random intercepts and slopes by participant, and random intercepts by item.

```
model_m_full <- M_Acc ~ Pronoun + (1 + Pronoun|SubjID) + (1 + Pronoun|Name)

model_m <- buildmer(model_m_full, data=m,
                    family='binomial', direction=c('order'))
```

```
## Determining predictor order
```

```
## Fitting via glm: M_Acc ~ 1
```

```
## Currently evaluating LRT for: Pronoun

## Fitting via glm: M_Acc ~ 1 + Pronoun

## Updating formula: M_Acc ~ 1 + Pronoun

## Fitting via glm: M_Acc ~ 1 + Pronoun

## Currently evaluating LRT for: 1 | Name, 1 | SubjID

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 | Name)

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 | SubjID)

## Updating formula: M_Acc ~ 1 + Pronoun + (1 | SubjID)

## Currently evaluating LRT for: 1 | Name, Pronoun | SubjID

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 | SubjID) + (1 |
##     Name)

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 + Pronoun |
##     SubjID)

## Updating formula: M_Acc ~ 1 + Pronoun + (1 + Pronoun | SubjID)

## Currently evaluating LRT for: 1 | Name

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 + Pronoun |
##     SubjID) + (1 | Name)

## Updating formula: M_Acc ~ 1 + Pronoun + (1 + Pronoun | SubjID) + (1 |
##     Name)

## Currently evaluating LRT for: Pronoun | Name

## Fitting via glmer, with ML: M_Acc ~ 1 + Pronoun + (1 + Pronoun |
##     SubjID) + (1 + Pronoun | Name)

## Ending the ordering procedure due to having reached the maximal
##     feasible model - all higher models failed to converge. The types of
##     convergence failure are: lme4 reports not having converged (-1)
```

```
summary(model_m)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) (p-values based on Wald z-scores) [glmerMod]
##  Family: binomial  ( logit )
## Formula: M_Acc ~ 1 + Pronoun + (1 + Pronoun | SubjID) + (1 | Name)
##    Data: m
##
##      AIC      BIC   logLik deviance df.resid
##   1431.7   1482.8   -705.9   1411.7     1214
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4508 -0.6637  0.4549  0.5837  1.9896
##
## Random effects:
##  Groups Name                   Variance Std.Dev. Corr
##  SubjID (Intercept)            0.442707 0.66536
##         Pronounthey vs he+she  0.634479 0.79654  -0.42
##         Pronounhe vs she       0.193204 0.43955   0.12  0.41
##  Name   (Intercept)            0.008379 0.09154
## Number of obs: 1224, groups:  SubjID, 102; Name, 12
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|) Pr(>|t|)
## (Intercept)               0.7683     0.1044  7.3593    0.000 1.85e-13 ***
## Pronounthey vs he+she     1.6382     0.1717  9.5436    0.000  < 2e-16 ***
## Pronounhe vs she          0.1292     0.2167  0.5962    0.551    0.551
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Prnvh+
## Prnnthyvsh+ 0.021
## Prononhvssh 0.060  0.084
```

Convert to odds:

```
exp(0.7683) #intercept (mean)
```

```
## [1] 2.156098
```

```
exp(1.6382) #they/them vs. he/him + she/her
```

```
## [1] 5.145899
```

- The intercept is significant (p<.001), such that participants are 2.16 times more likely to answer correctly than incorrectly across all pronoun types.
- The contrast between they/them and he/him + she/her is significant (p<.001), such that participants are 5.15 times as likely to get he/him and she/her right than they/them.
- The contrast between he/him and she/her is not significant.

Production

Descriptive Stats

Mean accuracy, split by pronoun type. Accuracy for producing they/them is lower than accuracy for remembering they/them.

```
prop.table(table(p$Pronoun, p$P_Acc), margin=1)
```

```
##
##                      0         1
##   he/him    0.1691176 0.8308824
##   she/her   0.1421569 0.8578431
##   they/them 0.7058824 0.2941176
```

60% of participants produced they/them at least once.

```
they_p <- d %>% filter(P_Pronoun=="they/them") %>%
  summarize(n=n_distinct(SubjID))

they_p/(n_distinct(d$SubjID))
```

```
##           n
## 1 0.5980392
```

Model

Same model specifications as first model (memory accuracy). Here, the maximal model has random intercepts and slopes by participant, and no random effects by item.

```
model_p_full <- P_Acc ~ Pronoun + (1 + Pronoun|SubjID) + (1 + Pronoun|Name)

model_p <- buildmer(model_p_full, data=p,
                    family='binomial', direction=c('order'))
```

```
## Determining predictor order

## Fitting via glm: P_Acc ~ 1

## Currently evaluating LRT for: Pronoun

## Fitting via glm: P_Acc ~ 1 + Pronoun

## Updating formula: P_Acc ~ 1 + Pronoun

## Fitting via glm: P_Acc ~ 1 + Pronoun

## Currently evaluating LRT for: 1 | Name, 1 | SubjID
```

```
## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + (1 | Name)

## boundary (singular) fit: see help('isSingular')

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + (1 | SubjID)

## Updating formula: P_Acc ~ 1 + Pronoun + (1 | SubjID)

## Currently evaluating LRT for: 1 | Name, Pronoun | SubjID

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + (1 | SubjID) + (1 |
##     Name)

## boundary (singular) fit: see help('isSingular')

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + (1 + Pronoun |
##     SubjID)

## Updating formula: P_Acc ~ 1 + Pronoun + (1 + Pronoun | SubjID)

## Currently evaluating LRT for: 1 | Name

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + (1 + Pronoun |
##     SubjID) + (1 | Name)

## Ending the ordering procedure due to having reached the maximal
##     feasible model - all higher models failed to converge. The types of
##     convergence failure are: lme4 reports not having converged (-1)
```

```
summary(model_p)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) (p-values based on Wald z-scores) [glmerMod]
##  Family: binomial  ( logit )
## Formula: P_Acc ~ 1 + Pronoun + (1 + Pronoun | SubjID)
##    Data: p
##
##      AIC      BIC   logLik deviance df.resid
##   1048.6   1094.6   -515.3   1030.6     1215
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5806 -0.2653  0.1580  0.3570  2.0167
##
## Random effects:
##  Groups Name                   Variance Std.Dev. Corr
##  SubjID (Intercept)             1.0056  1.0028
##         Pronounthey vs he+she 12.3364  3.5123    0.66
##         Pronounhe vs she       0.3173  0.5633   -0.94 -0.87
## Number of obs: 1224, groups:  SubjID, 102
```

```
##
## Fixed effects:
##                        Estimate Std. Error z value Pr(>|z|) Pr(>|t|)
## (Intercept)              1.3299     0.1977  6.7268    0.000 1.73e-11 ***
## Pronounthey vs he+she    4.1418     0.4705  8.8028    0.000  < 2e-16 ***
## Pronounhe vs she        -0.1569     0.4717 -0.3326    0.739    0.739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Prnvh+
## Prnnthyvsh+  0.527
## Prononhvssh -0.238 -0.192
```

Convert to odds:

```
exp(1.3299) #intercept (mean)
```

```
## [1] 3.780665
```

```
exp(4.1418) #they/them vs. he/him + she/her
```

```
## [1] 62.91597
```

- The intercept is significant (p<.001), such that participants are 3.78 times more likely to answer correctly than incorrectly across all pronoun types.
- The contrast between they/them and he/him + she/her is significant (p<.001), such that participants are 62.92 times more likely to get he/him and she/her right than they/them.
- The contrast between he/him and she/her is not significant.

Memory Predicting Production

Descriptive Stats

Combining the two measures, there are 4 possible patterns: getting both right, getting both wrong, getting just memory right, and getting just production right.

```
mp_acc <- mp %>%
          mutate(BothRight=ifelse(M_Acc==1 & P_Acc==1, 1, 0)) %>%
          mutate(BothWrong=ifelse(M_Acc==0 & P_Acc==0, 1, 0)) %>%
          mutate(MemOnly=ifelse(M_Acc==1 & P_Acc==0, 1, 0)) %>%
          mutate(ProdOnly=ifelse(M_Acc==0 & P_Acc==1, 1, 0)) %>%
          pivot_longer(cols=c(BothRight, BothWrong, MemOnly, ProdOnly),
                       names_to="Combined_Accuracy") %>%
          group_by(Pronoun, Combined_Accuracy) %>%
          summarise(m=mean(value))
```

```
## `summarise()` has grouped output by 'Pronoun'. You can override using the
## `.groups` argument.
```

```
mp_acc
```

```
## # A tibble: 12 x 3
## # Groups:   Pronoun [3]
##    Pronoun   Combined_Accuracy      m
##    <fct>     <chr>              <dbl>
##  1 he/him    BothRight         0.662
##  2 he/him    BothWrong         0.0686
##  3 he/him    MemOnly           0.100
##  4 he/him    ProdOnly          0.169
##  5 she/her   BothRight         0.689
##  6 she/her   BothWrong         0.0564
##  7 she/her   MemOnly           0.0858
##  8 she/her   ProdOnly          0.169
##  9 they/them BothRight         0.211
## 10 they/them BothWrong         0.480
## 11 they/them MemOnly           0.225
## 12 they/them ProdOnly          0.0833
```

Model

Model predicting production accuracy with pronoun type and memory accuracy. Otherwise the same model specifications as the first two. The maximal model includes random intercepts by participant.

```
model_mp_full <- P_Acc ~ M_Acc * Pronoun +
                 (1 + Pronoun|SubjID) + (1 + Pronoun|Name)

model_mp <- buildmer(model_mp_full, data=mp, family='binomial',
                     direction=c('order'))
```

```
## Determining predictor order

## Fitting via glm: P_Acc ~ 1

## Currently evaluating LRT for: M_Acc, Pronoun

## Fitting via glm: P_Acc ~ 1 + M_Acc

## Fitting via glm: P_Acc ~ 1 + Pronoun

## Updating formula: P_Acc ~ 1 + Pronoun

## Currently evaluating LRT for: M_Acc

## Fitting via glm: P_Acc ~ 1 + Pronoun + M_Acc

## Updating formula: P_Acc ~ 1 + Pronoun + M_Acc

## Currently evaluating LRT for: M_Acc:Pronoun
```

```
## Fitting via glm: P_Acc ~ 1 + Pronoun + M_Acc + M_Acc:Pronoun

## Updating formula: P_Acc ~ 1 + Pronoun + M_Acc + M_Acc:Pronoun

## Fitting via glm: P_Acc ~ 1 + Pronoun + M_Acc + M_Acc:Pronoun

## Currently evaluating LRT for: 1 | Name, 1 | SubjID

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc
##     + (1 | Name)

## boundary (singular) fit: see help('isSingular')

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc
##     + (1 | SubjID)

## Updating formula: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc + (1 |
##     SubjID)

## Currently evaluating LRT for: 1 | Name, Pronoun | SubjID

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc
##     + (1 | SubjID) + (1 | Name)

## boundary (singular) fit: see help('isSingular')

## Fitting via glmer, with ML: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc
##     + (1 + Pronoun | SubjID)

## Ending the ordering procedure due to having reached the maximal
##     feasible model - all higher models failed to converge. The types of
##     convergence failure are: Singular fit lme4 reports not having
##     converged (-1)
```

```
summary(model_mp)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) (p-values based on Wald z-scores) [glmerMod]
##  Family: binomial  ( logit )
## Formula: P_Acc ~ 1 + Pronoun + M_Acc + Pronoun:M_Acc + (1 | SubjID)
##    Data: mp
##
##      AIC      BIC   logLik deviance df.resid
##   1124.1   1159.9   -555.0   1110.1     1217
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9247 -0.4107  0.3204  0.4044  4.1244
##
## Random effects:
```

```
##  Groups Name        Variance Std.Dev.
##  SubjID (Intercept) 0.3469   0.589
## Number of obs: 1224, groups:  SubjID, 102
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|) Pr(>|t|)
## (Intercept)                   0.1287     0.1454  0.8853    0.376   0.3760
## Pronounthey vs he+she         3.0163     0.2726 11.0659    0.000  < 2e-16 ***
## Pronounhe vs she              0.2549     0.3452  0.7385    0.460   0.4602
## M_Acc                         1.2452     0.1682  7.4048    0.000 1.31e-13 ***
## Pronounthey vs he+she:M_Acc  -0.8196     0.3364 -2.4365    0.015   0.0148 *
## Pronounhe vs she:M_Acc       -0.0641     0.4292 -0.1493    0.881   0.8813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Prnvh+ Prnnvs M_Acc  Pvh+:M
## Prnnthyvsh+  0.224
## Prononhvssh  0.061  0.061
## M_Acc       -0.724 -0.145 -0.045
## Prvh+sh:M_A -0.163 -0.780 -0.043  0.158
## Prnnvsh:M_A -0.050 -0.051 -0.813  0.054  0.049
```

Convert to odds:

```
exp(1.24520) #memory accuracy
```

```
## [1] 3.473629
```

```
exp(-0.81954) #they/them vs. he/him + she/her * memory accuracy
```

```
## [1] 0.4406343
```

- The effect of memory accuracy is significant ($p<.001$), such that participants are 3.47x more likely to get the production right if they got the memory right.
- Significant interaction between pronoun type (they/them vs. he/him + she/her) and memory accuracy ($p<.05$) (odds 0.44). The relative difficulty of they/them was attenuated when the participant had correctly remembered the character's pronoun during the memory phase of the task.