The Sound Symbolism Experiment

Sebastian Scott Engen

12/06/2017

## Abstract

This experiment investigated the relation if vowel/consonant sounds could be used to predict classification of figures of different shapes and sizes. The sound symbolism experiment showed that the vowel O, compared to the vowel I,served as a significant postive predictor of big and curved figures. And that the consonant K, compared to B, served as a significant negative predictor of curved figures.

## Introduction

* The arbitraryness of language
* The kiki-Bouba effect (comparing to the original experiment)

## Materials and Methods

The experiment included 35 participant, who were all presented with 32 scenarios. Each scenario included two figure and one name that could be chosen for the given figure using the arrowkeys on the keyboard.

# Analysis

We used a mixed effects model with shape (Jagged or Curved) and size (Big or Small) as the outcome variables, consonants and vowels as fixed effects, and participant ID as random effects. size ~ vowel + (1|id) size ~ consonant + (1|id) shape ~ vowel + (1|id) shape ~ consonant + (1|id)

## Loading required package: Matrix

## Loading required package: lme4

## Warning: package 'lme4' was built under R version 3.4.2

##   
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':  
##   
## lmer

## The following object is masked from 'package:stats':  
##   
## step

## Loading required package: lattice

##   
## Attaching package: 'lattice'

## The following object is masked from 'package:boot':  
##   
## melanoma

## Loading required package: ggplot2

## Loading required package: Formula

## Loading required package: maxLik

## Loading required package: miscTools

##   
## Please cite the 'maxLik' package as:  
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood estimation in R. Computational Statistics 26(3), 443-458. DOI 10.1007/s00180-010-0217-1.  
##   
## If you have questions, suggestions, or comments regarding the 'maxLik' package, please use a forum or 'tracker' at maxLik's R-Forge site:  
## https://r-forge.r-project.org/projects/maxlik/

## Warning: package 'MuMIn' was built under R version 3.4.2

##   
## Attaching package: 'MuMIn'

## The following object is masked from 'package:miscTools':  
##   
## coefTable

##   
## Attaching package: 'car'

## The following object is masked from 'package:boot':  
##   
## logit

## Loading required package: splines

## Loading required package: RcmdrMisc

## Loading required package: sandwich

## Loading required package: effects

##   
## Attaching package: 'effects'

## The following object is masked from 'package:car':  
##   
## Prestige

## The Commander GUI is launched only in interactive sessions

# In terms of size and shape, we'll determine the baseline for R with releveling. That means that we´ll make predictions about how much more likely it is get big shapes compared to small and curved shapes compared to jagged:  
kikibobo3$size <- relevel(kikibobo3$size, "sma")  
# Small will be the baseline  
  
kikibobo3$shape <- relevel(kikibobo3$shape, "jagged")  
#Jagged will be the baseline  
  
#Size-Vowel model  
model1 = glmer(size ~ vowel + (1|id), data = kikibobo3, family = "binomial")  
summary(model1)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: size ~ vowel + (1 | id)  
## Data: kikibobo3  
##   
## AIC BIC logLik deviance df.resid   
## 1542 1557 -768 1536 1117   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.1547 -0.9078 0.8660 0.8660 1.1016   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0 0   
## Number of obs: 1120, groups: id, 35  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.19346 0.08491 -2.278 0.0227 \*   
## vowelO 0.48114 0.12042 3.995 6.46e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## vowelO -0.705

"  
Fixed effects:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -0.19346 0.08491 -2.278 0.0227 \*   
vowelO 0.48114 0.12042 3.995 6.46e-05 \*\*\*  
"

## [1] "\nFixed effects:\n Estimate Std. Error z value Pr(>|z|) \n(Intercept) -0.19346 0.08491 -2.278 0.0227 \* \nvowelO 0.48114 0.12042 3.995 6.46e-05 \*\*\*\n"

# The vocal O  
#Odds  
exp(-0.19346+0.48114)

## [1] 1.333331

# odds are 1.333331  
# Odds are above 1 - so we can expect that there is a higher probability of choosing a big model with the vowel O  
# The exact Probabilities  
inv.logit(-0.19346+0.48114)

## [1] 0.5714281

# the probabilities are 57% of choosing a big model when the vowel is an O   
  
#Size-consonant model  
model1.2 = glmer(size ~ consonant + (1|id), data = kikibobo3, family = "binomial")  
summary(model1.2)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: size ~ consonant + (1 | id)  
## Data: kikibobo3  
##   
## AIC BIC logLik deviance df.resid   
## 1557.6 1572.7 -775.8 1551.6 1117   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.0438 -1.0036 0.9580 0.9964 0.9964   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0 0   
## Number of obs: 1120, groups: id, 35  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.007143 0.084516 0.084 0.933  
## consonantK 0.078624 0.119578 0.658 0.511  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## consonantK -0.707

"  
Fixed effects:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.007143 0.084516 0.084 0.933  
consonantK 0.078624 0.119578 0.658 0.511  
"

## [1] "\nFixed effects:\n Estimate Std. Error z value Pr(>|z|)\n(Intercept) 0.007143 0.084516 0.084 0.933\nconsonantK 0.078624 0.119578 0.658 0.511\n"

# The P-value isn't significant so we won't try to find the probability for consonants effect on size  
  
  
  
  
#Shape-Consonant model  
model2 <- glmer(shape ~ consonant + (1|id), data = kikibobo3, family = "binomial")  
summary(model2)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: shape ~ consonant + (1 | id)  
## Data: kikibobo3  
##   
## AIC BIC logLik deviance df.resid   
## 1174.7 1189.8 -584.4 1168.7 1117   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2124 -0.6159 0.4520 0.4520 1.6237   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0 0   
## Number of obs: 1120, groups: id, 35  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.5882 0.1126 14.11 <2e-16 \*\*\*  
## consonantK -2.5576 0.1471 -17.39 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## consonantK -0.765

"  
Fixed effects:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 1.5882 0.1126 14.11 <2e-16 \*\*\*  
consonantK -2.5576 0.1471 -17.39 <2e-16 \*\*\*  
"

## [1] "\nFixed effects:\n Estimate Std. Error z value Pr(>|z|) \n(Intercept) 1.5882 0.1126 14.11 <2e-16 \*\*\*\nconsonantK -2.5576 0.1471 -17.39 <2e-16 \*\*\*\n"

# The consonant K  
#odds  
exp(1.5882-2.5576)

## [1] 0.3793106

# odds are 0.3793106  
# Odds are below 1 - so we can expect that there is a lower probability of choosing a curved model with the consonant k  
#Probabilities  
inv.logit(1.5882-2.5576)

## [1] 0.2750001

# the probabilities are 28% of choosing a curved model when the consonant is an K  
  
#Shape-Vowel model  
model2.2 <- glmer(shape ~ vowel + (1|id), data = kikibobo3, family = "binomial")  
summary(model2.2)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: shape ~ vowel + (1 | id)  
## Data: kikibobo3  
##   
## AIC BIC logLik deviance df.resid   
## 1499.7 1514.8 -746.9 1493.7 1117   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.3735 -0.9078 0.7280 0.7280 1.1016   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0 0   
## Number of obs: 1120, groups: id, 35  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.19346 0.08491 -2.278 0.0227 \*   
## vowelO 0.82823 0.12287 6.741 1.58e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## vowelO -0.691

"  
Fixed effects:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -0.19346 0.08491 -2.278 0.0227 \*   
vowelO 0.82823 0.12287 6.741 1.58e-11 \*\*\*  
"

## [1] "\nFixed effects:\n Estimate Std. Error z value Pr(>|z|) \n(Intercept) -0.19346 0.08491 -2.278 0.0227 \* \nvowelO 0.82823 0.12287 6.741 1.58e-11 \*\*\*\n"

# The vocal O  
#Odds  
exp(-0.19346+ 0.82823)

## [1] 1.886588

# odds are 1.886588  
# Odds are above 1 - so we can expect that there is a higher probability of choosing a curved model with the vowel O  
#Probabilities  
inv.logit(-0.19346+ 0.82823)

## [1] 0.6535703

# the probabilities are 65% of choosing a curved model when the vowel is an O

# Results

The sound symbolism experiment showed that the vowel O served as a significant positive predictor of the size 'big': b = 0.48114 (SE = 0.12042), z = 3.995, p < .001

The sound symbolism experiment showed that the consonant K served as a significant negative predictor of the shape 'curved': b = -2.5576 (SE = 0.1471), z = -17.39, p < .001

The sound symbolism experiment showed that the vowel O served as a significant positive predictor of the shape 'curved': b = 0.82823 (SE = 0.12287), z = 6.741, p < .001

## Discussion

Comment on the fact that our data shows the complete opposite of what was expected.

Limitations: - The participants in the study could be biased since they knew what theoretical work it was trying to replicate. - Next to this an important factor is that the participants might have been influenced by the fact that the figures were drawn up with lines of different thickness. Aditionaly to this, some of the figures were also blurry, which might have influenced the choice of the participants - Another factor that hasn't been accounted for either is that the internal pronounciation, in the participants heads, might be different duo to dialect and accent. - To this it could also be taking into consideration that some participants just looked at the letters of the words - e.g. the jaggedness of a K compared to a B - and not the sound-difference. - A fifth factor that might have influenced the choices of the participants is that they we're limited to choose from two figures at a time, so even though vowel O might generally be recognised as being part of words for Big shapes - the participant could be presented with only small figures and a word with the vowel O, which forces them to choose a small shape, while if they only had the choice to also choose a big figure, they would have.