Overconfidence for picture cues in foreign language learning

Jason Geller

Last Knitted: 2023-09-18

### Vignette Setup:

knitr::opts\_chunk$set(echo = TRUE)  
  
# Set a random seed  
set.seed(3898934)  
  
# Libraries necessary for this vignette  
library(rio)  
library(flextable)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(psych)  
library(semanticprimeR)  
  
# Function for simulation  
item\_power <- function(data, # name of data frame  
 dv\_col, # name of DV column as a character  
 item\_col, # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 200,   
 sample\_increase = 5,  
 decile = .5){  
   
 DF <- cbind.data.frame(  
 "dv" = data[ , dv\_col],  
 "items" = data[ , item\_col]  
 )  
   
 # just in case  
 colnames(DF) <- c("dv", "items")  
   
 # figure out the "sufficiently narrow" ci value  
 SE <- tapply(DF$dv, DF$items, function (x) { sd(x)/sqrt(length(x)) })  
 cutoff <- quantile(SE, probs = decile)  
   
 # sequence of sample sizes to try  
 samplesize\_values <- seq(sample\_start, sample\_stop, sample\_increase)  
  
 # create a blank table for us to save the values in   
 sim\_table <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF$items)))  
  
 # make it a data frame  
 sim\_table <- as.data.frame(sim\_table)  
  
 # add a place for sample size values   
 sim\_table$sample\_size <- NA  
  
 # loop over sample sizes  
 for (i in 1:length(samplesize\_values)){  
   
 # temp that samples and summarizes  
 temp <- DF %>%   
 group\_by(items) %>%   
 sample\_n(samplesize\_values[i], replace = T) %>%   
 summarize(se = sd(dv)/sqrt(length(dv)))  
   
 # dv on items  
 colnames(sim\_table)[1:length(unique(DF$items))] <- temp$items  
 sim\_table[i, 1:length(unique(DF$items))] <- temp$se  
 sim\_table[i, "sample\_size"] <- samplesize\_values[i]  
   
 }  
  
 # figure out cut off  
 final\_sample <- sim\_table %>%   
 pivot\_longer(cols = -c(sample\_size)) %>%   
 rename(item = name, se = value) %>%   
 group\_by(sample\_size) %>%   
 summarize(percent\_below = sum(se <= cutoff)/length(unique(DF$items)))   
   
 return(list(  
 SE = SE,   
 cutoff = cutoff,   
 DF = DF,   
 sim\_table = sim\_table,   
 final\_sample = final\_sample  
 ))  
  
}

### Project/Data Title:

Overconfidence for picture cues in foreign language learning

Data provided by: Jason Geller

### Project/Data Description:

Previous research shows that participants are overconfident in their ability to learn foreign language vocabulary from pictures compared with English translations. The current study explored whether this tendency is due to processing fluency or beliefs about learning. Using self-paced study of Swahili words paired with either picture cues or English translation cues, 30 participants provided JOLs to each of the 42 English-Swahili word pairs from Carpenter and Olson’s (2012) Experiment 2.The English words were one-syllable nouns, ranging between three and six letters, with an average concreteness rating of 4.86 on a 5-point scale (SD = .16) (Brysbaert, Warriner, & Kuperman, 2014), and an average frequency of 106.52 per million (SD = 113.40) (Brysbaert & New, 2009).

### Methods Description:

Participants began the experiment with instructions informing them that they would be learning Swahili words paired with either pictures or English translations as cues. To illustrate each type of cue, they were given an example of an item (Train: Reli) that was not included among the 42 experimental items. They were informed that each pair of items (English-Swahili pairs or picture-Swahili pairs) would be presented one at a time, and they would have as much time as they needed to study it. Participants were encouraged to do their best to learn each pair, and to encourage full and meaningful processing of each, they were instructed to press the spacebar once they felt they had fully “digested” it. For each participant, 21 items were randomly selected to be presented as English-Swahili pairs, and 21 as picture-Swahili pairs. Participants saw each stimulus pair one at a time, in a unique random order with English-Swahili pairs and picture-Swahili pairs intermixed. Each pair was presented in the center of the computer screen and remained on screen until participants pressed the spacebar to move on to the next pair. After each of the 42 pairs was presented for self-paced study in this way, the same pairs were presented again for JOLs. During a JOL trial, each cue-target pair was presented on the screen and participants were asked to estimate—using a scale from 0% (definitely will NOT recall) to 100% (definitely will recall)—the likelihood of recalling the Swahili word from its cue (either the picture or English translation) after about 5 minutes. Participants entered a value between 0 and 100 and pressed the ENTER key to advance to the next item.

### Data Location:

Data can be found here: <https://osf.io/2byt9/>.

#read in data  
DF <- import("geller\_data.xlsx") %>%   
 select(Experiment, Subject, `CueType[1Word,2Pic]`, Stimulus, EncodeJOL)

## Warning: Expecting numeric in J2069 / R2069C10: got 'jico'

str(DF)

## 'data.frame': 2898 obs. of 5 variables:  
## $ Experiment : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Subject : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ CueType[1Word,2Pic]: num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Stimulus : chr "kidoto" "muhindi" "kiti" "jaluba" ...  
## $ EncodeJOL : num 1 1 1 1 1 1 1 1 1 1 ...

### Date Published:

No official publication, see citation below.

### Dataset Citation:

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. Quarterly Journal of Experimental Psychology, 73(2), 211–224. <https://doi.org/10.1177/1747021819879416>

### Citations:

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. Quarterly Journal of Experimental Psychology, 73(2), 211–224. <https://doi.org/10.1177/1747021819879416>

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(1), 92–101. <https://doi.org/10.1037/a0024828>

### Keywords:

Overconfidence, metacognition, processing fluency, analytic processing, foreign language learning

### Use License:

Open access with reference to original paper (Attribution-NonCommercial-ShareAlike CC BY-NC-SA)

### Geographic Description - City/State/Country of Participants:

Ames, Iowa

### Column Metadata:

metadata <- tibble::tribble(  
 ~Variable.Name, ~Variable.Description, ~`Type (numeric,.character,.logical,.etc.)`,  
 "Experiment", "Experiment 1 (1) or 2 (2) ONLY USE 1", NA,  
 "Subject", "Subject ID", "Numeric",  
 "CueType", "Whether participant was presented with word translation (1) or word with picture (2)", "Numeric",  
 "Stimulus", "Swahili words presented on each trail", "Character",  
 "EncodeJOL", "JOL (1-100) 1=not likely to recall 100=very likely to recall", "Numeric"  
 )  
  
flextable(metadata) %>% autofit()

| Variable.Name | Variable.Description | Type (numeric,.character,.logical,.etc.) |
| --- | --- | --- |
| Experiment | Experiment 1 (1) or 2 (2) ONLY USE 1 |  |
| Subject | Subject ID | Numeric |
| CueType | Whether participant was presented with word translation (1) or word with picture (2) | Numeric |
| Stimulus | Swahili words presented on each trail | Character |
| EncodeJOL | JOL (1-100) 1=not likely to recall 100=very likely to recall | Numeric |

### AIPE Analysis:

#### Stopping Rule

DF <- DF %>%   
 filter(Experiment == 1) %>%  
 filter(!is.na(EncodeJOL))  
  
# Function for simulation  
var1 <- item\_power(data = DF, # name of data frame  
 dv\_col = "EncodeJOL", # name of DV column as a character  
 item\_col = "Stimulus", # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 100,   
 sample\_increase = 5,  
 decile = .4)

What the usual standard error for the data that could be considered for our stopping rule using the 40% decile?

# individual SEs  
var1$SE

## andiko bao bunduki chaka chapeo chimbule daraja dawati   
## 5.569863 5.421327 6.467708 5.812192 5.232406 5.964306 5.006286 4.295635   
## dubu duwara farasi fia fupa gari geli jaja   
## 6.825573 3.857377 6.126705 5.192728 6.954286 6.092600 5.496488 7.424850   
## jaluba jicho jiti jumba juya kanisa kelb kidoto   
## 4.803243 5.649133 6.229624 5.405121 3.783108 6.245590 6.311649 4.324860   
## kipira kitanda kiti maliki mapwa mkono mlango muhindi   
## 5.895944 5.515605 6.404627 6.614947 6.634286 7.609701 5.745819 5.471669   
## muundi papatiko pua rinda riza safina samaki simu   
## 6.108910 6.681585 6.309811 5.378363 6.045881 6.548968 5.384973 5.490716   
## ufunguo wardi   
## 6.229707 5.739385

var1$cutoff

## 40%   
## 5.601571

Using our 40% decile as a guide, we find that 5.602 is our target standard error for an accurately measured item.

#### Minimum Sample Size

To estimate minimum sample size, we should figure out what number of participants it would take to achieve 80%, 85%, 90%, and 95% of the SEs for items below our critical score of 5.602?

cutoff <- calculate\_cutoff(population = DF,   
 grouping\_items = "Stimulus",  
 score = "EncodeJOL",  
 minimum = as.numeric(min(DF$EncodeJOL)),  
 maximum = as.numeric(max(DF$EncodeJOL)))  
# showing how this is the same as the person calculated version versus semanticprimeR's function  
cutoff$cutoff

## 40%   
## 5.601571

final\_table <- calculate\_correction(  
 proportion\_summary = var1$final\_sample,  
 pilot\_sample\_size = length(unique(DF$Subject)),  
 proportion\_variability = cutoff$prop\_var  
 )  
  
flextable(final\_table) %>%   
 autofit()

| percent\_below | sample\_size | corrected\_sample\_size |
| --- | --- | --- |
| 88.09524 | 35 | 31.48876 |
| 88.09524 | 35 | 31.48876 |
| 100.00000 | 40 | 39.49014 |
| 100.00000 | 40 | 39.49014 |

Our minimum sample size is (*n* = 31 as the minimum at 80%). We could consider using 90% (*n* = 39) or 95% (*n* = 39).

#### Maximum Sample Size

While there are many considerations for maximum sample size (time, effort, resources), if we consider a higher value just for estimation sake, we could use *n* = 39 at nearly 100%.

#### Final Sample Size

In any estimate for sample size, you should also consider the potential for missing data and/or unusable data due to any other exclusion criteria in your study (i.e., attention checks, speeding, getting the answer right, etc.). In this study, these values may be influenced by the pictures/word split in the study.