Replication of the Online Visual Crowding Experiment Using the Virtual Chinrest by Li et al.(2020, Scientific Reports)

Wanjing Anya Ma wanjingm@stanford.edu (mailto:wanjingm@stanford.edu)

December 06, 2021

Introduction

Online experiments enable psychophysical researchers to study larger and more diverse participant samples than they could have in the laboratory but leave challenges in controlling participants' viewing distance and the presentation of visual stimuli. The Virtual Chinrest, introduced and validated in Li et al. (2020), was designed to tackle these challenges by an estimation of the viewing distance through measuring the participant's blind spot location. The authors included two main parts: (1) two validation experiments of the Virtual Chinrest in the lab with participants being instructed to sit at varying viewing distances, and (2) a successful online replication of a lab-based study on visual crowding using the Virtual Chinrest.

Justification for choice of study

I chose to replicate this study for three reasons. (1) As a new student of psychophysical research, I am interested in learning a whole validation pipeline from a simple controlled lab environment to an uncontrolled online experiment. Due to the time constraint, I will replicate the uncontrolled online experiment (Study 4) only. (2) Replicating this study will enhance my technical skills in using jsPsych and Javascript. (3) More importantly, the second part of the study involved implementing an adaptive algorithm to present stimuli, which matches my research interest.

Anticipated challenges

The major challenge I anticipate is the time constraint. Implementing an adaptive visual crowding experiment might be time-consuming if the original source code is not available. In addition, compared with the original online study that was available online for 15 months to collect data, I will not be able to replicate the whole findings in the last experiment (e.g., the difference in the viewing distance between people with and without dyslexia).

Links

Project repository (on Github): [AnyaWMa] https://github.com/psych251/li2020.git (https://github.com/psych251/li2020.git)

Original paper (as hosted in your repo): https://github.com/psych251/li2020/blob/main/original_paper/Li_et_al-2020-Scientific_Reports.pdf (https://github.com/psych251/li2020/blob/main/original_paper/Li_et_al-2020-Scientific_Reports.pdf)

Methods

Power analysis

The power analysis of the original study.

```
library(esc)

esc_mean_se(grp1m = 1.623,   # mean of group 1
            grp1se = 0.592,  # standard error of group 1
            grp1n = 742,    # sample in group 1
            grp2m = 2.584,    # mean of group 2
            grp2se = 1.162,  # standard error of group 2
            grp2n = 280,    # sainmple in group 2
            es.type = "d") # convert to SMD; use "g" for Hedges' g
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.0563
##  Standard Error:   0.0701
##        Variance:   0.0049
##        Lower CI:  -0.1938
##        Upper CI:   0.0812
##          Weight: 203.2237
```

```
## two-sample
power.welch.t.test(delta = 0.961, sd1 = 0.592, sd2 = 1.163, power = 0.8)
```

```
##
##      Two-sample Welch t test power calculation
##
##               n = 15.8446
##           delta = 0.961
##             sd1 = 0.592
##             sd2 = 1.163
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Planned sample

In order to meet 80% power, I plan to have 16 participants for each condition, so there will be 32 participants in total. Only neurotypical, young adults (age range: 18-30) will be recruited in this replication study.

Experimental design

“The online experiment was launched on the volunteer-based online experiment platform LabintheWild and advertised with the slogan “How accurate is your peripheral vision?” on the site itself as well as on social media. Experimental design. During each experimental session, we first presented the Virtual Chinrest experiment and used the results to calculate individual’s viewing distance and to calibrate the stimuli’s size and locations. Instead of creating stimuli (demonstrated in Fig. 5) using MATLAB, we created the stimuli as SVG on HTMLs and manipulated the stimuli using JavaScript. All the elements were created in a container with a width of 900 pixels on the webpage. In the blind spot test, the dot was drawn in red with a diameter of 30 pixels, and the fixation square was drawn in black with a side length of 30 pixels (Fig. 1b). Replicating the original crowding study30 in the unit of visual degrees, stimuli comprised four flankers — open circles with 1° diameter and a target — an open circle with a gap (target; an arc with reflex central angle of 330°). All stimuli were black and displayed on a white background (Fig. 5). Two conditions of target eccentricity (the center-to-center distance between the fixation mark at the center of the webpage and the target) were 4° and 6°. In each crowding experiment session, each participant was randomly assigned one target eccentricity, and the target eccentricity was fixed with the starting target-flanker distance being set as 1.3 times greater than half the eccentricity (3.9° for 6° eccentricity; 2.6° for 4° eccentricity). During each crowding experiment session, the subsequent target-flanker distances (25 trials/steps in total) were controlled by the 1-up 3-down staircase procedure implemented in JavaScript [https://github.com/hadrienj/StaircaseJS (https://github.com/hadrienj/StaircaseJS)]. On a given trial, the fixation mark was displayed first and remained on the webpage for the entire session. After 500 ms of fixation onset, the stimuli were displayed either to the left or the right of the fixation for 150 ms. Only the fixation remained on the webpage until the participant submitted a response by using the arrow keys on the keyboard to indicate the direction (up or down) of the target gap. No feedback was provided during the experiment. There was a 500 ms blank between a participant’s response and the beginning of the next trial. The visual crowding, defined as the minimal center-to-center distance between a target and the flankers (in degrees), was used to quantify the crowding effects when participants could report the target identity at certain accuracy. Since we are using a 1-up 3-down staircase procedure, participants should be able to correctly report the target identity 79.4% of times."

Procedure

> “The experiment began with a brief overview of the study, an informed consent form approved by the University of Washington Institutional Review Board, and a voluntary demographic questionnaire, followed by the card task and the blind spot test with 5 trials to calculate participants’ viewing distances. Participants were then presented the instruction of the crowding tasks and a practice session with 5 trials. The main experiment was split into two blocks (two independent staircases, 25 trials each), and each was followed by another blind spot task with 3 trials. After the last blind spot test, participants were then given the opportunity to report on any technical difficulties, and to provide any other general comments or questions. The final page showed their personalized “crowding effect” in comparison to others. The entire study took 10–12 minutes to complete."

Analysis plan

> “We deployed the online study in two stages, where we added more granular data log at the second stage, such as the percentage correctness of the experiment and the results of each individual trial. Therefore, the analysis of visual crowding effects (Fig. 7a,b) was performed on the data of 793 participants from the second stage, the results in Table 2 was based on a subset of 570 participants who have explicitly reported whether they have dyslexia and/or other related impairments, while the results of the viewing distances from the three blind (Fig. 7c–e) spot tests were reported from all 1153 participants. We checked for data normality by both the visual inspection of histograms and the Shapiro-Wilk normality tests before each analysis. We then conducted parametric (e.g. the Welch’s two sample t-test) and non-parametric (e.g. Mann-Whitney U test) analysis accordingly. In the linear mixed-effects regression models, t-tests (p-values) were calculated using Satterthwaite approximations for the degrees of freedom. The data analysis of all four experiments was performed in R, with the help of multiple packages.”

Differences from original study

Here are three major differences between this replication study and the original study.

Experiment implementation

The original study implemented the Virtual Chinrest and visual crowding experiment using JavaScript. Only part of the code used jsPsych library. In my replication, I will refer the high level idea of the original source code, but I decide to use the existing Virtual Chinrest jsPsych plugin and build the entire visual crowding experiment through jsPsych. The code, free from the original lab's coding framework, should be more concise and more accessible for future replication studies.

Sampling

The original experiment was deployed online for 15 months and completed 1198 times, and after data exclusion, the valid data contained 793 participants. The large and diverse sample size allowed the researchers to study how different covariates (eccentricity, gender, age, and with/without dyslexia) relate to the effect visual crowding. However, due to the time and financial constraints, this current replication tests the effect of eccentricity among 18-30 year-old, neurotypical participants only. The sample size is about 40 (20 for each condition).

Data collection

The original study recruited participants on the volunteer-based experiment platform LabintheWild, while this replication project used Prolific, a crowdsourcing platform recruit participants with pay. Using Prolific may lead to higher quality data because participants are paid and it doesn't need to rely on participants' self-report whether they have completed the study before.

Data from the original study was written into the first author's MySQL database. However, I will use Pavlovia to host the experiment and store data directly into my GitLab.

Methods addendum

Pilot A with known participants from friends and labmates

The Pilot A data included a total of 6 participants (3 for each eccentricity condition).

1. Virtual Chinrest: two participants reported that the red dot did not go far enough to the left if the participant was too far away from the screen.
2. Instructions: One participants reported that there were a lot of instructions on the peripheral vision task page. She suggested "I would break it up in bullet points or present 1 and 2 on separate pages, reading long lines of instructions is not encouraging."
3. Practice Trials: One participant said she would have a better idea of the experiment if she could have more practice trials.

To fix the problem 1, I need to check the jsPsych plugin code to adjust the movement of the red dot in the blind spot measurement trial. However, I will not resolve problem 2 and 3 to be faithful to the original study, although they are good suggestions.

Pilot B with unknown participants from Prolific

1. Prescreening Criteria:
    a. Age 18-30
    b. Language related disorder: No
    c. Literacy Difficulty: No
    d. Mild Cognitive Impairment/Dementia: No
    e. Autism Spectrum Disorder: No
    f. Normal Vision or Corrected Vision: Yes
    g. Exclude participants from previous visual crowding studies: Yes h: Prolific Approval Rate: 95%-100% i: Confidential agreement: Yes
2. I recruited 2 participants through Prolific for each condition for pilot B. 3 out of 4 participants completed the study. The 4th pariticipant didn't proceed to the last page of the experiment, so his/her results were not saved.

3. Compared with the result of Pilot A, the Pilot B data got different number of participants between 2 conditions. This was expected for 2 reasons.

   First, in an crowd sourcing environment, participants fall under the exclusion criteria will not excluded in the final analyses.

   Second, the original paper specified that the display eccentricity should not exceeds 225mm after adjusting the stimuli based on participant's screen size and viewing distance. If exceeds under the the eccentricity 6° condition， participants will be automatically changed to receive the eccentricity 4° condition. If the eccentricity 4° condition still exceeds, then the participant will be tested on 225mm mode. With this experimental constraint, I will expect more participants who are tested under eccentricity 4° condition although I counterbalance the two conditions at the beginning. It is noted that eccentricity 6° had 742 non-dyslexic entries while eccentricity 4° had 280 non-dyslexic entries in the original study.

Exclusion criteria

In my final analysis, I plan to exclude participants who spent less than 4 minutes in the entire study because they are likely to rush the experiment and input off-task responses.

Alternative Plan: If I exclude more than 3 participants in either condition, I will run the experiment with 10 more participants for each condition.

Actual sample

The prescreening criteria in the final sampling were consistent with the criteria used in Pilot B (see above). Among the n = 43 participants recruited to participate in the final study, n = 37 participants completed the experiment and saved their responses successfully to GitLab. All 37 participants spent more than 4 minutes in the study, so none was excluded in the final analysis. The completion rate of this study was 86.0%.

In the visual crowding experiment, each valid participant completed two sessions, so there were 74 entries in total. As expected and described above, some participants who were initially assigned to take eccentricity 6° condition had to be modified to take the eccentricity 4° condition or even smaller eccentricity deg if their adjusted display-relative eccentricity exceeded 225mm. Therefore, only in the visual crowding analysis, 50 entries under eccentricity 4° and 22 entries under eccentricity 6° were kept, while 2 entries with eccentricity lower than 4° were removed.

Differences from pre-data collection methods plan

None.

Results

The analysis were followed by [the public R code]
https://github.com/QishengLi/virtual_chinrest/blob/master/analysis/Exp_4_online_experiment_analysis.
(https://github.com/QishengLi/virtual_chinrest/blob/master/analysis/Exp_4_online_experiment_analysis
provided by the first author.

Data preparation

Load relevant libraries and functions

```r
#### Load Relevant Libraries and Functions
library(data.table)
library(readr)
library(printr)
library(readr)
library(psych)
library(tidyr)
library(ggplot2)
library(GGally)
library(lmerTest)
library(dplyr)
library(plyr)
library(stringr)
library(rstatix)
library(hablar)
library(cowplot)
```

Helper functions

```r
#sem: calculate standard error from the mean
sem <- function(x) sd(x)/sqrt(length(x))

#conversion between radian and degree
rad2deg <- function(rad) {(rad * 180) / (pi)}
deg2rad <- function(deg) {(deg * pi) / (180)}
```

Import data

```r
### Data Preparation
myfiles = list.files(path = "../data/final-protected" , pattern= "*.csv",
full.names=TRUE)
df = ldply(myfiles, read_csv)
```

Data exclusion / filtering

```r
filtered_df <-
  df  %>%
  drop_na(participantID) %>%
  filter(correctRate > 0.0)
```

```r
visual_crowding <- subset(filtered_df, select= c(participantID, eccentrici
ty_deg, critical_spacing_deg, crowding_session, correctRate, age))
#visual_crowding
```

remove the participant entry which was automatically adjusted to a lower eccentricity_deg (neither 4 deg or 6 deg)

```r
visual_crowding_revised <-
  visual_crowding  %>%
  filter(eccentricity_deg >= 4.00)

visual_crowding <- visual_crowding_revised

#check number of entries available for the confirmatory analysis
visual_crowding %>% group_by(eccentricity_deg) %>% tally()
```

eccentricity_deg	n
4	50

prepare a dataframe to analyze the viewing distance.

```
distance_df <-
  df  %>%
  filter(grepl("end_trial", trialType))

viewing_distance_df <- subset(distance_df, select= c(participantID,viewDis
tance_1, viewDistance_2, viewDistance_3 ))

#viewing_distance_df
```

```
viewing_distance_reshape_df <- reshape(data= viewing_distance_df, idvar="p
articipantID",varying = c("viewDistance_1", "viewDistance_2", "viewDistanc
e_3"),v.name=c("viewDistance"),direction="long", times = c(1,2,3))

#viewing_distance_reshape_df
```

Confirmatory analysis

Figure 7a:

Visual crowding in two eccentricity (4 vs. 6) conditions

The major inference from the original paper I wanted to justify was: > “The average visual crowding effects were significantly different between target eccentricity of 4° (mean = 1.62°) and 6° (mean = 2.58°)” (Figure 7a).

```
#Mean
aggregate(critical_spacing_deg~eccentricity_deg, data = visual_crowding, m
ean)
```

eccentricity_deg	critical_spacing_deg
4	1.585200
6	1.893182

```
#SD
aggregate(critical_spacing_deg~eccentricity_deg, data = visual_crowding, s
d)
```

eccentricity_deg	critical_spacing_deg
4	0.6788232
6	0.8989167

```
#Standard error of the mean (sem)
aggregate(critical_spacing_deg~eccentricity_deg, data = visual_crowding, s
em)
```

eccentricity_deg	critical_spacing_deg
4	0.0960001
6	0.1916497

Non-parametric Test:

```
#Mean, SD, Standard error of the mean (sem)
visual_crowding_by <- visual_crowding %>%
  dplyr::group_by(eccentricity_deg) %>%
  dplyr::summarise(mean = mean(critical_spacing_deg),sd = sd(critical_spac
ing_deg), sem = sd(critical_spacing_deg)/sqrt(length(critical_spacing_deg)
))

#wilcoxon rank sum test
wilcox.test(critical_spacing_deg~eccentricity_deg,data=visual_crowding)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  critical_spacing_deg by eccentricity_deg
## W = 390, p-value = 0.04967
## alternative hypothesis: true location shift is not equal to 0
```

Parametric Test:

```
# Save the data in two different vector
 eccen4 <- visual_crowding %>%
  filter(eccentricity_deg == 4.0) %>%
  pull(critical_spacing_deg)
eccen6 <- visual_crowding %>%
  filter(eccentricity_deg == 6.0) %>%
  pull(critical_spacing_deg)
# Compute t-test
res <- t.test(eccen4, eccen6)
res
```

```
##
##  Welch Two Sample t-test
##
## data:  eccen4 and eccen6
## t = -1.4368, df = 31.997, p-value = 0.1605
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7445986  0.1286349
## sample estimates:
## mean of x mean of y
##  1.585200  1.893182
```

Calculate the effect size:

```
visual_crowding%>% cohens_d(critical_spacing_deg ~ eccentricity_deg , var.
equal = FALSE)
```

.y.	group1	group2	effsize	n1	n2	magnitude
critical_spacing_deg	4	6	-0.3866647	50	22	small

Reproduce Figure 7a:

```
cr_bars <- aggregate(critical_spacing_deg~eccentricity_deg, data = visual_
crowding, mean)
cr_stderr <- visual_crowding %>%
  dplyr::group_by(eccentricity_deg) %>%
  dplyr::summarise(sem = sd(critical_spacing_deg)/sqrt(length(critical_spa
cing_deg)))
cr_bars <- merge(cr_bars, cr_stderr, by=c('eccentricity_deg'))
```

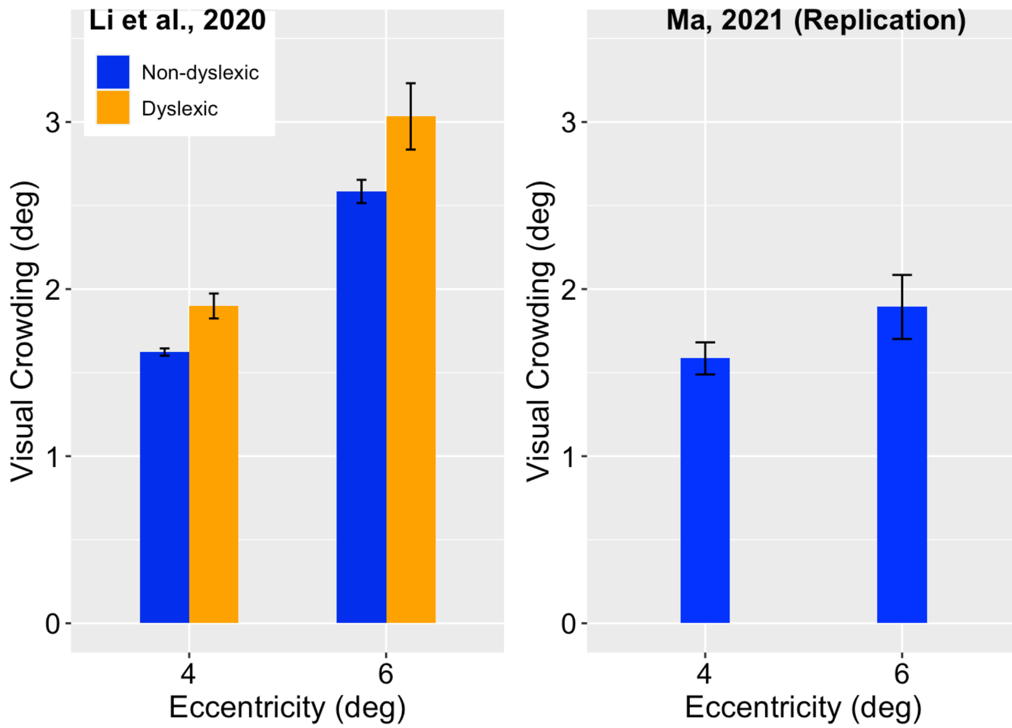
```
replication_plot_7a <- ggplot(data=cr_bars, aes(x=as.factor(eccentricity_d
eg), y=critical_spacing_deg)) +
  geom_bar(stat="identity", position=position_dodge(), width = 0.25, fill
= "blue")  +
  geom_errorbar(aes(ymin=critical_spacing_deg-sem, ymax=critical_spacing_d
eg+sem), width=.1, position=position_dodge()) +
  labs(x = "Eccentricity (deg)", y = "Visual Crowding (deg)") +
  theme(axis.text=element_text(size=14, color='black'), axis.title=element
_text(size=15), legend.text=element_text(size=10))+ coord_cartesian(ylim=c
(0,3.5))
```

```
#Draw Fig. 7a

original_cr_bars <- aggregate(critical_spacing_deg~eccentricity_deg+dyslex
ia, data = original_visual_crowding_dyslexia, mean)
original_cr_stderr <- aggregate(critical_spacing_deg~dyslexia+eccentricity
_deg, data = original_visual_crowding_dyslexia, sem)
names(original_cr_stderr)[3]<-paste("original_stderr")
original_cr_bars <- merge(original_cr_bars, original_cr_stderr, by=c('dysl
exia','eccentricity_deg'))

# reproduce original 7a:
original_plot_7a <- ggplot(data=original_cr_bars, aes(x=as.factor(eccentri
city_deg), y=critical_spacing_deg, fill=as.factor(dyslexia))) + geom_bar(s
tat="identity", position=position_dodge(), width = 0.5) + geom_errorbar(ae
s(ymin=critical_spacing_deg-original_stderr, ymax=critical_spacing_deg+ori
ginal_stderr), width=.1, position=position_dodge(.5)) +
  labs(x = "Eccentricity (deg)", y = "Visual Crowding (deg)") +
  scale_fill_manual("", labels = c("Non-dyslexic", "Dyslexic"), values=c("
blue2","orange")) +
  theme(axis.text=element_text(size=14, color='black'), axis.title=element
_text(size=15), legend.text=element_text(size=10)) +
  theme(legend.position = c(0.25, 0.9)) + coord_cartesian(ylim=c(0,3.5))
```

```
#Side-by-side graph:
# compare 7a
plot_grid(original_plot_7a, replication_plot_7a, labels = c('Li et al., 20
20', 'Ma, 2021 (Replication)'))
```




```
original_visual_crowding_no_dyslexia <- original_visual_crowding_dyslexia
%>%
  filter(dyslexia == 0)
```

Other analyses covered in the original study

Figure 7b:

Histgram of participant accuracy in the visual crowding experiment

```
# reproduce original 7b: histgram of participant accuracy
original_visual_crowding_accuracy <- aggregate(correctRate~uuid, data = or
iginal_visual_crowding_no_dyslexia, mean)

original_plot_7b <- ggplot(original_visual_crowding_accuracy, aes(x=correc
tRate)) + geom_histogram(aes(y=..density..), binwidth = 0.02, alpha=.8) +
  geom_vline(aes(xintercept=mean(correctRate)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5)) +
  labs(title="", x="Percentage Correctness", y="Density") +
  theme_grey(base_size = 15) +   theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10))

#summary: mean, min, max
summary(original_visual_crowding_accuracy$correctRate)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.42	0.84	0.88	0.8558513	0.9	1

```
visual_crowding_accuracy <- aggregate(correctRate~participantID, data = vi
sual_crowding, mean)

#histgram of participant accuracy
replication_plot_7b <- ggplot(visual_crowding_accuracy, aes(x=correctRate)
) + geom_histogram(aes(y=..density..), binwidth = 0.02, alpha=.8) +
  geom_vline(aes(xintercept=mean(correctRate)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5)) +
  labs(title="", x="Percentage Correctness", y="Density") +
  theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10)) + coord_cartesian(ylim=
c(0,10.0))

#summary: mean, min, max
summary(visual_crowding_accuracy$correctRate)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.46	0.86	0.9	0.8713514	0.92	0.96

```
# compare 7b
plot_grid(original_plot_7b, replication_plot_7b, labels = c('Li et al., 20
20', 'Ma, 2021 (Replication)'))
```

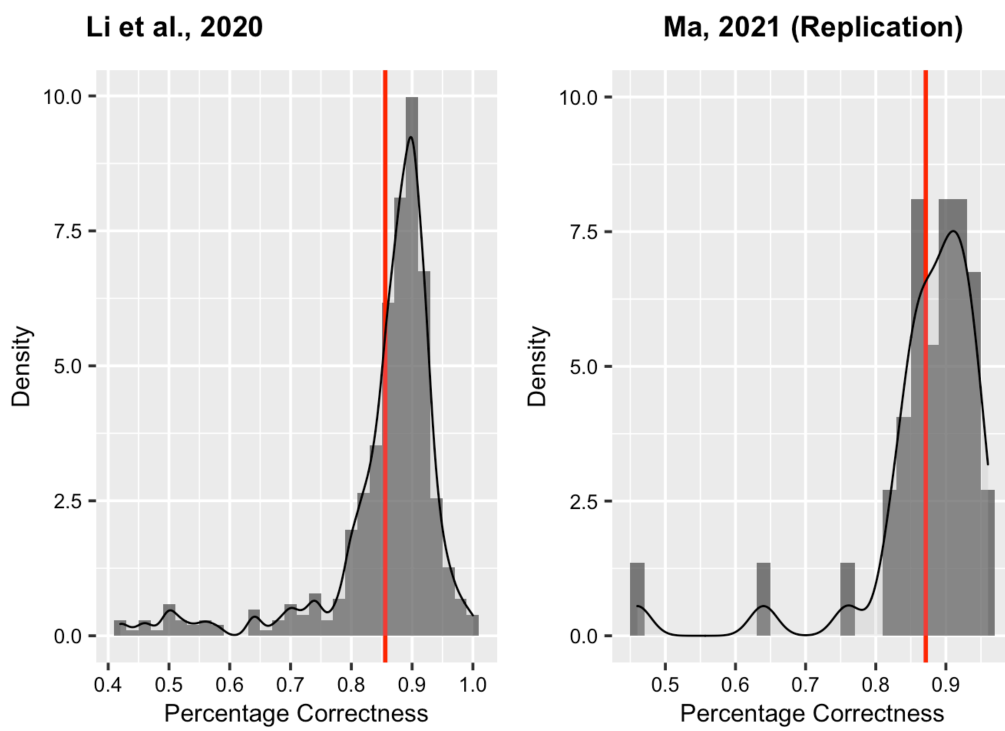


Figure 7c:

Distribution of calculated viewing distances

```
# reproduce original 7c
original_viewing_distance_mean <- aggregate(viewDistance~uuid, data = orig
inal_viewing_distance, mean)

original_plot_7c <- ggplot(original_viewing_distance_mean, aes(x=viewDista
nce)) + geom_histogram(aes(y=..density..), binwidth = 2, alpha=.8) +
  geom_vline(aes(xintercept=mean(viewDistance)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  labs(title="", x="Calculated Viewing Distance (cm)", y="Density") +
  theme_grey(base_size = 15) +   theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10)) +  coord_cartesian(ylim
=c(0,0.08))
```

```
#data summary
viewing_distance_mean <- aggregate(viewDistance~participantID, data = view
ing_distance_reshape_df, mean)
summary(viewing_distance_reshape_df$viewDistance) #mean, min, max
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
23.022	44.454	51.952	51.2875	59.8615	69.176

```
sd(viewing_distance_mean$viewDistance) # SD = 8.9
```

```
## [1] 8.516002
```

```
replication_plot_7c <- ggplot(viewing_distance_mean, aes(x=viewDistance))
+ geom_histogram(aes(y=..density..), binwidth = 2, alpha=.8) +
  geom_vline(aes(xintercept=mean(viewDistance)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  labs(title="", x="Calculated Viewing Distance (cm)", y="Density") +
  theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10)) + coord_cartesian(ylim=
c(0,0.08))
```

```
# compare 7c
plot_grid(original_plot_7c, replication_plot_7c, labels = c('Li et al., 20
20', 'Ma, 2021 (Replication)'))
```

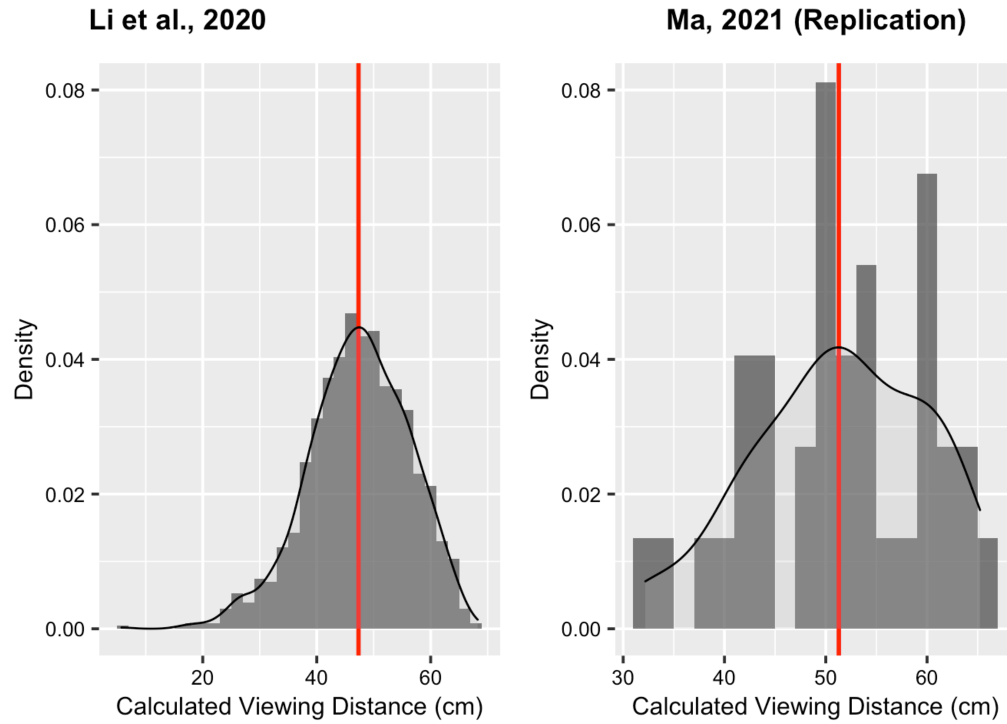


Figure 7d:

Within-subject standard deviation across three blind spot tests: mean, min, max of SD of calculated viewing distances.

```
# reproduce original 7d
original_viewing_distance_sd <- aggregate(viewDistance~uuid, data = origin
al_viewing_distance, sd)

original_plot_7d <- ggplot(original_viewing_distance_sd, aes(x=viewDistanc
e)) + geom_histogram(aes(y=..density..), binwidth = 0.5, alpha=.8) +
  geom_vline(aes(xintercept=mean(viewDistance)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  labs(title="", x="SD of Calculated Viewing Distance (cm)", y="Density")+
  theme_grey(base_size = 15) +
    theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10)) +  coord_cartesian(ylim
=c(0,0.5))
```

```
viewing_distance_sd <- aggregate(viewDistance~participantID, data = viewin
g_distance_reshape_df, sd)

#draw Fig. 7d
replication_plot_7d <- ggplot(viewing_distance_sd, aes(x=viewDistance)) +
geom_histogram(aes(y=..density..), binwidth = 0.5, alpha=.8) +
  geom_vline(aes(xintercept=mean(viewDistance)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  labs(title="", x="SD of Calculated Viewing Distance (cm)", y="Density")+
  theme_grey(base_size = 15) +
  theme(axis.text=element_text(size=10, color='black'), axis.title=element
_text(size=12), legend.text=element_text(size=10)) + coord_cartesian(ylim=
c(0,0.5))
```

```
# compare 7d
plot_grid(original_plot_7d, replication_plot_7d, labels = c('Li et al., 20
20', 'Ma, 2021 (Replication)'))
```

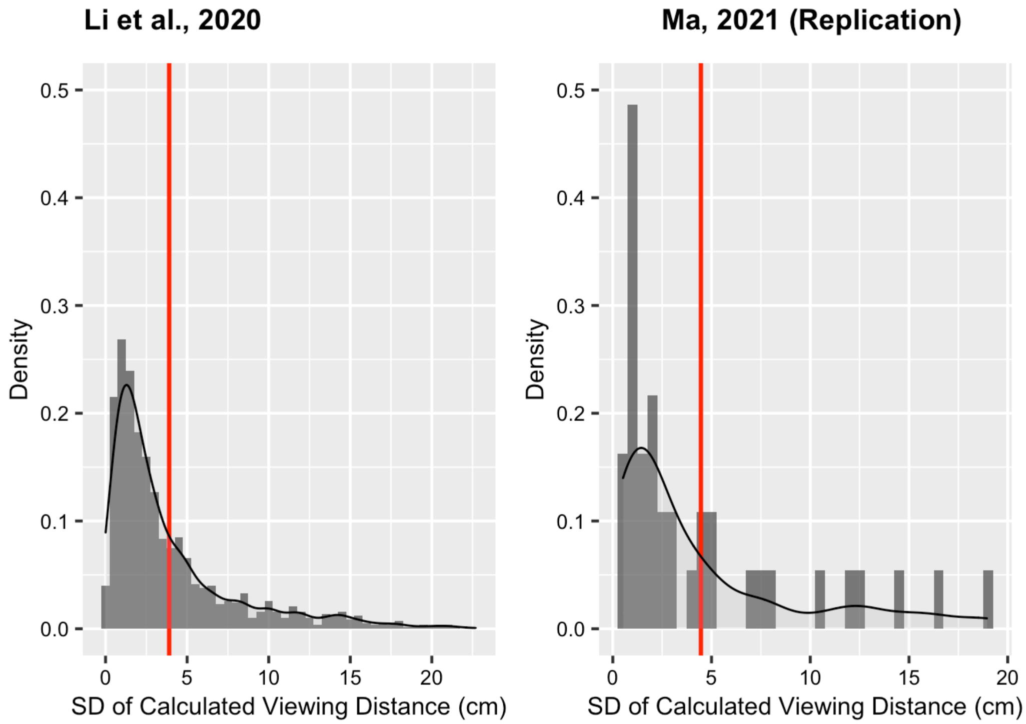


Figure 7e:

Within-subject viewing distance ICC and pairwise correlation.

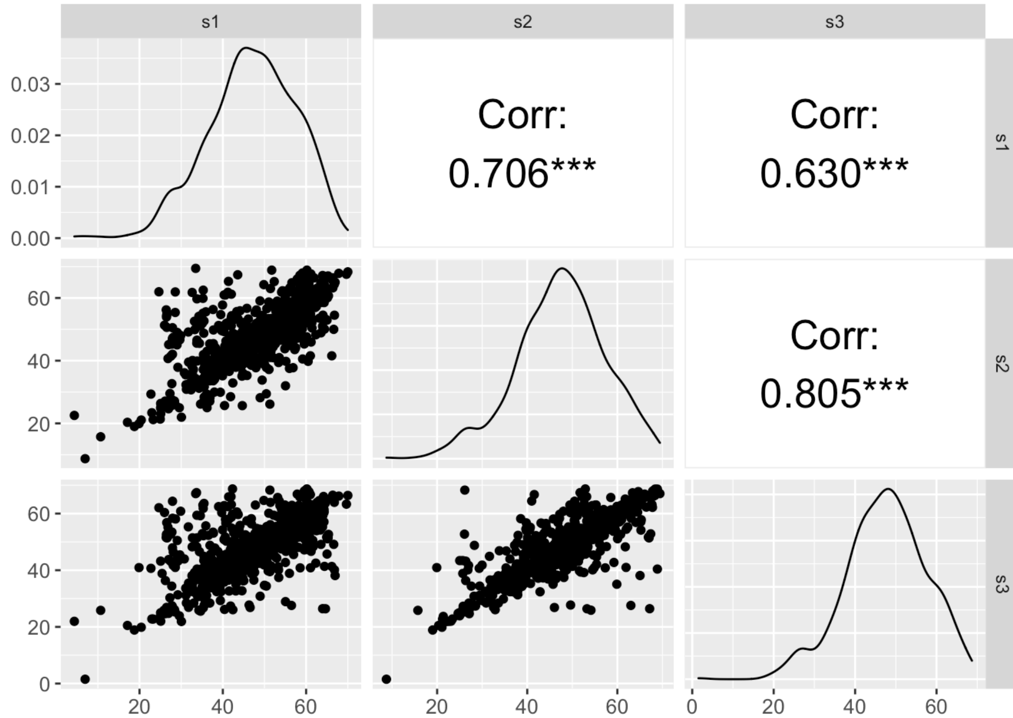
Original 7e (Li e al., 2020):

```
# reproduce original 7e
#transfer the df from long to wide
original_viewing_distance_wide <- spread(original_viewing_distance_by_sess
ion, session, viewDistance)
names(original_viewing_distance_wide)[2:4] <- c('s1','s2','s3')

#calculate original ICC
ICC(original_viewing_distance_wide[,c(2,3,4)],missing=TRUE,alpha=.05)
```

```
## Call: ICC(x = original_viewing_distance_wide[, c(2, 3, 4)], missing = T
RUE,
##     alpha = 0.05)
##
## Intraclass correlation coefficients
##                          type  ICC   F df1  df2        p lower bound
## Single_raters_absolute   ICC1 0.71 8.4 804 1610 2.2e-282        0.69
## Single_random_raters     ICC2 0.71 8.4 804 1608 2.2e-282        0.69
## Single_fixed_raters      ICC3 0.71 8.4 804 1608 2.2e-282        0.69
## Average_raters_absolute ICC1k 0.88 8.4 804 1610 2.2e-282        0.87
## Average_random_raters   ICC2k 0.88 8.4 804 1608 2.2e-282        0.87
## Average_fixed_raters    ICC3k 0.88 8.4 804 1608 2.2e-282        0.87
##                         upper bound
## Single_raters_absolute         0.73
## Single_random_raters           0.73
## Single_fixed_raters            0.73
## Average_raters_absolute        0.89
## Average_random_raters          0.89
## Average_fixed_raters           0.89
##
##  Number of subjects = 805     Number of Judges =  3
## See the help file for a discussion of the other 4 McGraw and Wong estim
ates,
```

```
#draw Fig. 7e
original_plot_7e <- ggpairs(original_viewing_distance_wide, columns = 2:4,
upper = list(continuous = wrap("cor", size=7, color='black'))) +
  theme(axis.text = element_text(size = 10))
original_plot_7e
```



Replication 7e (Ma, 2021):

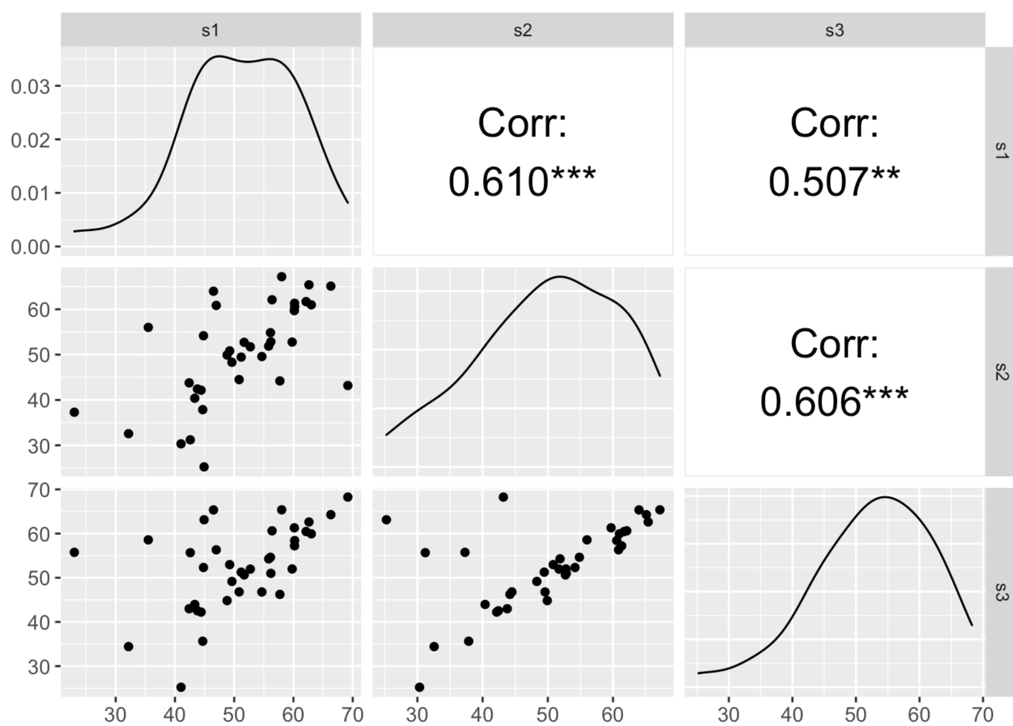
```
#transfer the df from long to wide
viewing_distance_wide <- spread(viewing_distance_reshape_df, time, viewDis
tance)
names(viewing_distance_wide)[2:4] <- c('s1','s2','s3')

#calculate ICC
ICC(viewing_distance_wide[,c(2,3,4)],missing=TRUE,alpha=.05)
```

```
## Call: ICC(x = viewing_distance_wide[, c(2, 3, 4)], missing = TRUE,
##     alpha = 0.05)
##
## Intraclass correlation coefficients
##                          type  ICC F df1 df2       p lower bound upper
bound
## Single_raters_absolute   ICC1 0.57 5  36  74 2.3e-09        0.42
0.71
## Single_random_raters     ICC2 0.57 5  36  72 2.8e-09        0.42
0.71
## Single_fixed_raters      ICC3 0.57 5  36  72 2.8e-09        0.42
0.71
## Average_raters_absolute ICC1k 0.80 5  36  74 2.3e-09        0.69
0.88
## Average_random_raters   ICC2k 0.80 5  36  72 2.8e-09        0.69
0.88
## Average_fixed_raters    ICC3k 0.80 5  36  72 2.8e-09        0.69
0.88
##
##  Number of subjects = 37     Number of Judges =  3
## See the help file for a discussion of the other 4 McGraw and Wong estim
ates,
```

```
#draw Fig. 7e
replication_plot_7e <- ggpairs(viewing_distance_wide, columns = 2:4, upper
= list(continuous = wrap("cor", size=7, color='black'))) +
  theme(axis.text = element_text(size = 10))

replication_plot_7e
```



Exploratory analyses

The original study has a broader age range (7-71), while my study only recruited young adults (18-30). The original study and their references showed older adults exhibited a larger crowding effect, particularly at greater eccentricity (6 deg). The following analyses will plot and compare 1) the distribution of age and 2) the the impact of age on the visual crowding effect between the original and my replication study.

Distribution of age

```
original_visual_crowding_age <- aggregate(age~uuid, data = original_visual
_crowding_no_dyslexia, mean)

#histgram of participant accuracy
original_plot_age <- ggplot(original_visual_crowding_age, aes(x=age)) + ge
om_histogram(aes(y=..density..), binwidth = 1.0, alpha=.8) +
  geom_vline(aes(xintercept=mean(age)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5)) +
  labs(title="", x="age", y="Density") +
  theme_grey(base_size = 15)

#summary: mean, min, max
summary(original_visual_crowding_age$age)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7	19	22	27.28376	33	71

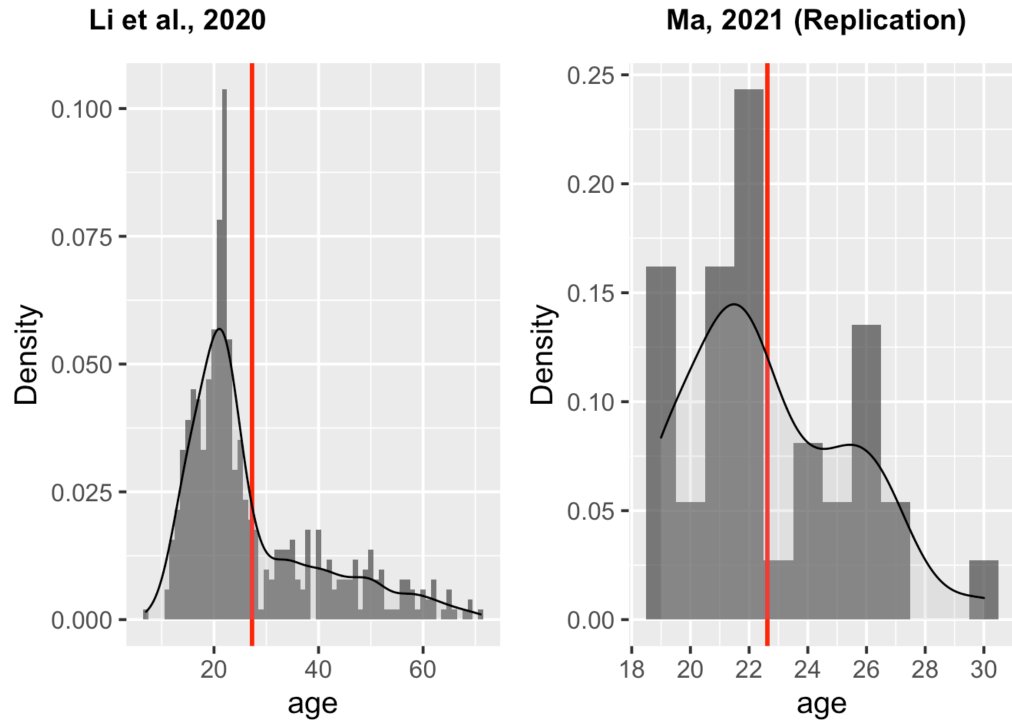
```
visual_crowding_age <- aggregate(age~participantID, data = visual_crowding
, mean)

#histgram of participant accuracy
replication_plot_age <- ggplot(visual_crowding_age, aes(x=age)) + geom_his
togram(aes(y=..density..), binwidth = 1.0, alpha=.8) +
  geom_vline(aes(xintercept=mean(age)), color="red", size=1) +
  geom_density(alpha=.2, fill="grey") +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 5)) +
  labs(title="", x="age", y="Density") +
  theme_grey(base_size = 15)

#summary: mean, min, max
summary(visual_crowding_age$age)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
19	21	22	22.62162	25	30

```
# compare age distribution
plot_grid(original_plot_age, replication_plot_age, labels = c('Li et al.,
2020', 'Ma, 2021 (Replication)'))
```



Relationship between age and visual crowding effect

Original (Li et al., 2020)

```
original_visual_crowding_no_dyslexia <- original_visual_crowding_no_dyslex
ia %>%
  convert(chr(eccentricity_deg))

ggplot(original_visual_crowding_no_dyslexia, aes(x=age, y=critical_spacing
_deg, color=eccentricity_deg,shape = eccentricity_deg)) +
  geom_point() +
  geom_smooth(method=lm)
```

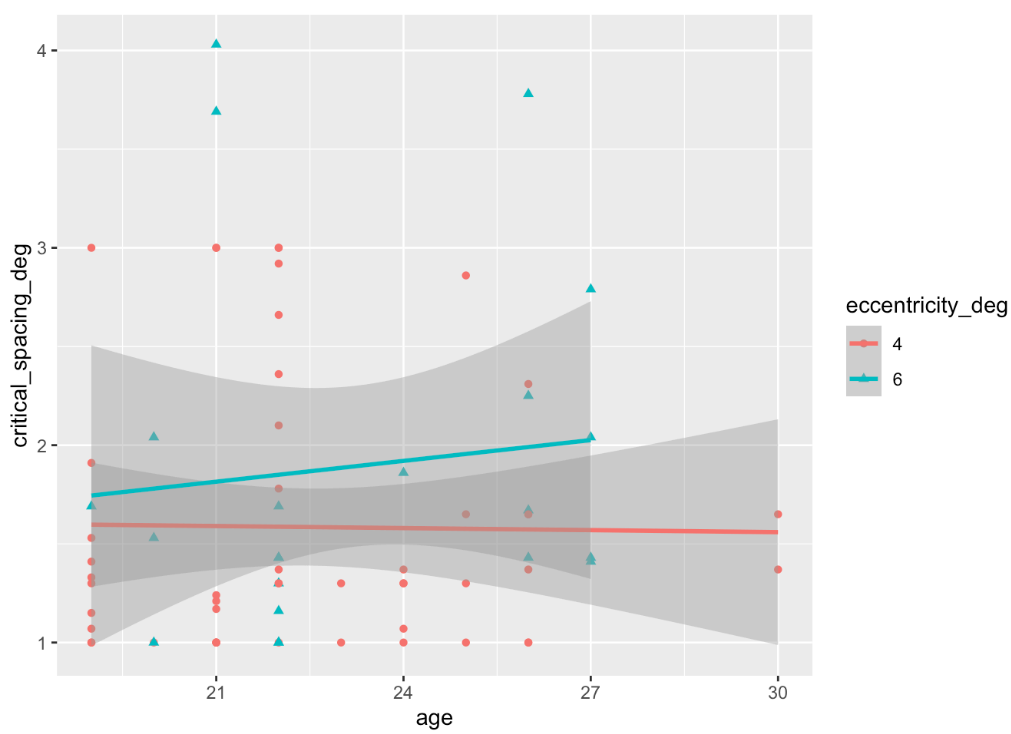
```
## `geom_smooth()` using formula 'y ~ x'
```

Replication (Ma, 2021)

```
convert_visual_crowding <- visual_crowding %>%
  convert(chr(eccentricity_deg))

ggplot(convert_visual_crowding, aes(x=age, y=critical_spacing_deg, color=e
ccentricity_deg,shape = eccentricity_deg)) +
  geom_point() +
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Discussion

Summary of replication attempt

Open the discussion section with a paragraph summarizing the primary result from the confirmatory analysis and the assessment of whether it replicated, partially replicated, or failed to replicate the original result.

The major inference from the original paper I wanted to justify was:

> "The average visual crowding effects were significantly different between target eccentricity of 4° (mean = 1.62°) and 6° (mean = 2.58°)" (Figure 7a).

Through a non-parametric Mann-Whitney U test, I replicated the significant difference ($W = 390$, p-value < 0.05) between the target eccentricity of 4° (mean = 1.59, sd = 0.68) and 6° (mean = 1.89, sd = 0.90). However, I failed to replicate the difference between the two conditions is using a parametric Welch's two sample t-test ($t(31.997) = -1.4$, $p = 0.16$, $d = -0.38$). Moreover, the mean (mean = 1.89) of visual crowding effect under eccentricity 6° in my replication was much lower than the mean (mean = 2.58) in the original study.

Besides of the confirmatory analysis, I was able to replicate the percentage correctness in the visual crowding experiment (Figure 7b), the viewing distance distribution (Figure 7c; Figure 7d), the and the strong pairwise correlations within-subject viewing distance (Figure 7e).

I conclude that my project partially replicated the original findings.

Commentary

Assessment of Replication

There are three potential reasons that could explain my replicate could not fully replicate the original findings.

1. Sample size: I had a very small sample (37 valid participants) compared with the original study.

2. Platform: Instead of collecting data "in the wild" in the original study, I recruited participants from Prolific who might be more skilled at participating psychophysics experiments. The analysis above shows that the average correctness rate (mean = 0.87) in my replication was slightly higher than the correctness rate in the original study (mean = 0.85).

3. Age: The original study has a broader age range (7-71), while my study only recruited young adults (18-30). The original study and their references showed age could older adults exhibited a larger crowding effect, particularly at greater eccentricity (6 deg). However, my exploratory analyses showed there was not a statistical difference in terms of age between the original study (mean = 27, median = 22) and my replication (mean = 23, median = 22). And the scatterplot "Relationship between age and visual crowding effect" demonstrates that, under eccentricity 6° condition, younger participants aged around 22 in the original sample showed crowding effect above 2.5°, still much higher than my replication 1.89°. Therefore, I cannot conclude that age was the primary reason for the failure to replicate.

4. Experimental Measurement: The Virtual Chinrest I used was a jsPsych plugin built by the jsPsych team who referred to the original authors' code. This new plugin, however, has not been validated by the physical chinrest in the lab. Therefore, it is likely that the plugin didn't calculate the viewing distance as consistently as the original built-in Virtual

Chinrest, which further resulted in the inconsistent display of the stimuli in my replication. I am interested in conducting validation tasks about the plugin Virtual Chinrest using the physical chinrest in my lab.

Reflections

1. The original authors did a wonderful job on organizing and sharing the data, analysis code, and source code.
2. Pilot A involving expert group and non-expert group could give you feedback from different perspectives.
3. jsPsych is very user-friendly. Building a framework-free paradigm is key to future replications.
4. Replicating a psychophysics experiment was challenging but very fun.

Acknowledgment

I want to express my gratitude to my instructor Dr. Mike Frank and my TAs, Sarah Wu, Julie Cachia, and Effie Li, who guided me and offered me detailed support throughout the project. I would also like to thank my adviser Dr. Jason Yeatman who introduced the original paper to me and encouraged me to learn jsPsych. Lastly, I want to thank my friends and labmates who participated in my Pilot A study and gave constructive feedback.

References

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. Scientific reports, 10(1), 1-11.