Computational reproduction of Freud et al. (2022)

Charlie Homewood

Last complied: 26/01/2023

Summary

This document runs through each analysis reported in Freud et al. (2022) in turn, ending with a discussion about the performance of the computational reproduction and issues faced throughout.

Freud et al. (2022) Outline

Freud et al. (2022) aimed to determine if real-world exposure to masked faces during the COVID-19 pandemic improved the ability of adults to recognise masked faces. 1768 participants were assessed across 6 time points (May 2020, Sept 2020, Jan 2021, May 2021, Sept 2021 & Jan 2022). This study comprised of a cross-sectional design using the Cambridge Face Memory Test (CFMT) across the 6 time points, and a longitudinal design over 12 months.

Supplementary materials

The authors provided the datasets used in this study on OSF (https://osf.io/tq92h), but despite claiming to provide analysis code (p. 1637) they did not do so. Some additional information regarding the contents of each dataset was provided (https://osf.io/tq92h/wiki/home/) although this was not obvious to find.

Contents

Read in datasets 4

Inspect datasets 4

Computational Reproduction 5
- Cross Sectional Analysis 5
  - ANOVA (time points, orientation, gender and group) 5
    - Original paper reported: 6
    - Computational reproduction success: 7
    - Issues faced: 7
  - Bayesian ANOVA 7
    - Original paper reported: 8
    - Computational reproduction success: 8
    - Issues faced: 9
  - Two-sample Kolmogorov-Smirnov test 9
    - Original paper reported: 9
    - Computational reproduction success: 9
    - Issues faced: 9
  - ANOVA for reaction times (RTs) 9
    - Original paper reported: 10
    - Computational reproduction success: 10
    - Issues faced: 10
  - Bayesian ANOVA for RTs 10
    - Original paper reported: 11
    - Computational reproduction success: 11
    - Issues faced: 11
- Longitudinal Analysis 11
  - Repeated measures ANOVA (group, orientation, time) 11
    - Original paper reported: 13
    - Computational reproduction success: 13
    - Issues faced: 13
  - ANOVA RTs 13
    - Original paper reported: 15
    - Computational reproduction success: 15
    - Issues faced: 15
  - Correlation (Jan 2021, Jan 2022) 15
    - Original paper reported: 16
    - Computational reproduction success: 16
    - Issues faced: 16
- The EMF Questionnaire 17
  - EMF ANOVA 17
    - Original paper reported: 18
    - Computational reproduction success: 18
    - Issues faced: 18
  - Repeated measures ANOVA (group, orientation, work location) 18
    - Original paper reported: 19
    - Computational reproduction success: 19
    - Issues faced: 19
  - Bayesian ANOVA (masked upright) 19
    - Original paper reported: 19
    - Computational reproduction success: 19
    - Issues faced: 20

Correlation (Masked and Masked-Unmasked) 20
Original paper reported: 22
Computational reproduction success: 22
Issues faced: 22
CFMT: within-subject design 22
Repeated measures ANOVA (mask status as within-subjects variable) 22
Original paper reported: 23
Computational paper reported: 23
Issues faced: 23
ANOVA RTs 23
Original paper reported: 24
Computational reproduction success: 24
Issues faced: 25
Correlation (mask effect [accuracy and RTs], longitudinal EMF scores) 25
Original paper reported: 25
Computational reproduction success: 25
Issues faced: 25
ANOVA occupations (health & retail, students & unemployed) 25
Original paper reported: 25
Computational reproduction success: 25
Issues faced: 25
ANOVA (workplace, mask status) 25
Original paper reported: 25
Computational reproduction success: 26
Issues faced: 26
GFMT: masked faces 26
Correlation (face-perception performance [d' & RTs], EMF scores) 26
Original paper reported: 27
Computational reproduction success: 28
Issues faced: 28

Conclusion **28**

References **29**

Read in datasets

```r
##### LOAD IN DATASETS #####

# 01_all_timepoints_trimmed.csv

df_01 <- readr::read_csv(
  "Data/Original datasets/01_all_timepoints_trimmed.csv",
  show_col_types = FALSE
  )

# 02_CFMTsimple_Jan2022vsJan2021.csv

df_02 <- readr::read_csv(
  "Data/Original datasets/02_CFMTsimple_Jan2022vsJan2021.csv",
  show_col_types = FALSE
  )

# 03_CFMTwithinSubectSep2021.csv
# name of df_03 csv file has a typo which could make reproduction more difficult

df_03 <- readr::read_csv(
  "Data/Original datasets/03_CFMTwithinSubectSep2021.csv",
  show_col_types = FALSE
  )

# 04_GFMTOct2021_masked_version.csv

df_04 <- readr::read_csv(
  "Data/Original datasets/04_GFMTOct2021_masked_version.csv",
  show_col_types = FALSE
  )
```

Inspect datasets

```r
##### INSPECT DATASETS #####

# df_01
## By inspecting df_01 it appears that column '...1' refers to the subject ID
## for the whole dataset, whilst 'Unnamed: 0' refers to the ID for each ## time
## period. Therefore, rename '...1' to 'subjectID' and remove 'Unnamed: 0'.

df_01 <- df_01 %>%
  select(-'Unnamed: 0') %>%
  rename(subjectID = '...1')

# df_02
## By inspecting df_02 it appears the same situation as df_01 arises. Therefore,
## rename '...1' to 'subjectID' and remove 'Unnamed: 0'. df_02 has a
## pre-existing column called 'subjectID' which lists csv files. Therefore,
## remove this pre-existing subjectID column.
```

```
df_02 <- df_02 %>%
  select(-'Unnamed: 0', -'subjectID') %>%
  rename(subjectID = '...1')

# df_03
## By inspecting df_03 it appears that column '...1' also is the subject ID.
## Therefore, rename '...1' to 'subjectID'

df_03 <- df_03 %>%
  rename(subjectID = '...1')

# df_04
## By inspecting df_03 it appears that column '...1' also is the subject ID.
## Therefore, rename '...1' to 'subjectID'

df_04 <- df_04 %>%
  rename(subjectID = '...1')
```

Computational Reproduction

The results section (p. 1640-1646) lists 6 main categories of the analysis: "Cross-sectional analysis", "longitudinal analysis", "Individual differences with masked faces", "The EMF questionnaire", "CFMT: within-subjects design" & "GFMT: masked faces". All categories, with the exception of "individual differences with masked faces" has reported analyses. Computational reproduction of these categories with analyses will therefore be attempted in turn.

Cross Sectional Analysis

ANOVA (time points, orientation, gender and group)

```
##### CS ANOVA data setup #####

# Results section reports that observations made in Sept 2020 were excluded
# from this analysis

cs_aov_data <- df_01 %>%
 dplyr::filter(
    time_as_number != 2
  ) # excludes sept 2020 datapoints

# now pivot the data from wide to long

cs_aov_data <- cs_aov_data %>%
  tidyr::pivot_longer(
    cols = c("upright", "inverted"), # selects columns to be pivoted
    names_to = "orientation",        # name of the long formatted column
    values_to = "CFMT"               # values of the long formatted column
  ) # pivots data from wide to long format

# ANOVA requires subjectID, orientation, CFMT, time, gender and group

cs_aov_data <- cs_aov_data %>%
  select(
```

```r
    subjectID, time, group, gender, orientation, CFMT
  ) # selects required columns for this ANOVA
```

```r
##### CS ANOVA #####

# original results report a partial eta squared which is not default output for aov_ez

cs_aov <- afex::aov_ez(
  id = "subjectID",                         # independent variable
  dv = "CFMT",                              # dependent variable
  within = "orientation",                   # within-subject variables
  between = c("group", "gender", "time"), # between-subject variables
  data = cs_aov_data,                       # data
  anova_table = 'pes'                       # we want partial eta squared ('pes'),
                                            # is set to 'ges' by default
) # the ANOVA object

knitr::kable(
  nice(cs_aov),
  caption = "Cross-sectional ANOVA output (time points, orientation, gender and group)."
  ) %>% # creates nice ANOVA table output
  kable_styling(
    latex_options = "HOLD_position"
    )
```

Table 1: Cross-sectional ANOVA output (time points, orientation, gender and group).

Effect	df	MSE	F	pes	p.value
group	1, 1570	88.92	214.16 ***	.120	<.001
gender	1, 1570	88.92	66.41 ***	.041	<.001
time	4, 1570	88.92	3.07 *	.008	.016
group:gender	1, 1570	88.92	1.90	.001	.169
group:time	4, 1570	88.92	1.09	.003	.358
gender:time	4, 1570	88.92	2.72 *	.007	.028
group:gender:time	4, 1570	88.92	0.39	<.001	.815
orientation	1, 1570	30.60	2846.12 ***	.644	<.001
group:orientation	1, 1570	30.60	232.24 ***	.129	<.001
gender:orientation	1, 1570	30.60	2.95 +	.002	.086
time:orientation	4, 1570	30.60	0.09	<.001	.986
group:gender:orientation	1, 1570	30.60	0.89	<.001	.347
group:time:orientation	4, 1570	30.60	0.36	<.001	.840
gender:time:orientation	4, 1570	30.60	0.39	<.001	.819
group:gender:time:orientation	4, 1570	30.60	0.82	.002	.510

Original paper reported: "The ANOVA demonstrated main effects of group (nonmasked > masked), F(1, 1570) = 214, p < .001, ηp^2 = .12, orientation (upright > inverted), F(1, 1570) = 2846, p < .001, ηp^2 = .64, and gender (female > male), F(1, 1570) = 66.4, p < .001, ηp^2 = .04. An interaction was found between orientation and group, with a reduced inversion effect for masked faces, F(1, 1570) = 232.2, p < .001, ηp^2 = .129, that was consistent across time points (three-way interaction: F < 1)… …although we found a small effect for time point, F(4, 1570) = 3.065, p = .016, ηp^2 = .008 (with a higher average for the May 2021 and January 2022 samples), there was no interaction between time point and group, F(4, 1570) = 1.095, p = .357, ηp^2 = .003." (p. 1640).

Computational reproduction success: Results from this ANOVA were successfully reproduced in whole (see table 1).

Issues faced: There were difficulties in interpreting what the dependent variable was for the model used in this ANOVA. It was not obvious to me that the ‘upright’ and ‘inverted’ columns in the “01_all_timepoints_trimmed.csv” dataset needed to be pivoted into a long format (the orientation column) in order to create the orientation variable for the ANOVA.

Furthermore, this ANOVA was a repeated measures ANOVA which was not specified in the paper itself, this made running the ANOVA difficult as it being a repeated measures ANOVA had to be inferred.

Bayesian ANOVA

```
##### CS Bayesian ANOVA (upright-masked) #####

# need to:
## pivot upright and inverted to make CFMT score
## filter for the upright-masked faces condition
## select time and CFMT as variables for the Bayesian ANOVA

cs_bayes_up <- df_01 %>%
  pivot_longer(
    cols = c("upright", "inverted"), # selects columns to be pivoted
    names_to = "orientation",        # name of the long formatted column
    values_to = "CFMT"               # values of the long formatted column
  ) %>% # pivots data from wide to long format
  filter(
    orientation == "upright", # filters upright orientation
    group == "mask"           # filters mask group
  ) %>%
  select(
    time, CFMT # selects time and CFMT
  )

cs_bayes_up$time <- as.factor(cs_bayes_up$time) # converts the time IV into factor
```

```
##### CS Bayesian ANOVA (inverted-masked) #####

# need to:
## pivot upright and inverted to make CFMT score
## filter for the inverted-masked faces condition
## select time and CFMT as variables for the Bayesian ANOVA

cs_bayes_inv <- df_01 %>%
  pivot_longer(
    cols = c("upright", "inverted"), # selects columns to be pivoted
    names_to = "orientation",        # name of the long formatted column
    values_to = "CFMT"               # values of the long formatted column
  ) %>% # pivots data from wide to long format
  filter(
    orientation == "inverted", # filters inverted orientation
    group == "mask"            # filters mask group
  ) %>%
  select(
```

```
    time, CFMT # selects time and CFMT
  )

cs_bayes_inv$time <- as.factor(cs_bayes_inv$time) # converts the time IV into factor
```

```
##### CS Bayesian ANOVA output tables #####

# Bayesian ANOVA for upright-masked condition

anovaBF(
  formula = CFMT ~ time, # linear model formula
  data = cs_bayes_up     # the cs_bayes_up data
) %>%
  kable(
    caption = "Cross-sectional Bayesian ANOVA of time and CFMT score for the upright-masked condition."
) %>%
  kable_styling(
    latex_options = "HOLD_position"
    )
```

Table 2: Cross-sectional Bayesian ANOVA of time and CFMT score for the upright-masked condition.

	bf	error	time	code
time	0.0028349	0.0014701	Thu Jan 26 09:49:13 2023	1364106fa1b

```
# Bayesian ANOVA for inverted-masked condition

anovaBF(
  formula = CFMT ~ time, # linear model formula
  data = cs_bayes_inv    # the cs_bayes_inv data
) %>%
  kable(
    caption = "Cross-sectional Bayesian ANOVA of time and CFMT score for the inverted-masked condition.
) %>%
  kable_styling(
    latex_options = "HOLD_position"
    )
```

Table 3: Cross-sectional Bayesian ANOVA of time and CFMT score for the inverted-masked condition.

	bf	error	time	code
time	0.0029023	0.0014219	Thu Jan 26 09:49:13 2023	1364c8b267f

Original paper reported: "Next, to further evaluate the consistency of recognition performance of masked faces across time points, we focused our analysis on the upright-masked-faces condition. We conducted a Bayesian ANOVA with time point (six levels) on the accuracy scores… …Bayesian ANOVA decisively supported the null hypothesis (i.e., no difference between the six time points; BF10 = 0.003); the null hypothesis was 332 times more likely than the alternative hypothesis. Similar results were observed when the Bayesian ANOVA was employed on the masked inverted faces (BF10 = 0.003)." (p. 1641).

Computational reproduction success: Results from these Bayesian ANOVAs were almost entirely successfully reproduced (see tables 2 & 3). Discrepancies appear to be due to rounding errors.

Issues faced: The authors originally used JASP to perform these Bayesian ANOVAs. However, I used R to reproduce these analyses. As such, there were slight discrepancies between the reported results and the reproduced results. It was judged that these discrepancies were due to rounding errors rather than being indicative of genuine discrepancies between the original analyses and the reproduced analyses.

Two-sample Kolmogorov-Smirnov test

Original paper reported: "Two-sample Kolmogorov-Smirnov tests were employed across possible combinations for the masked and nonmasked conditions and confirmed that the distributions were not different from each other at the different time points (Ds < 0.2, p $> .2$)." (p. 1641-1642).

Computational reproduction success: This KS test was unsuccessfully reproduced.

Issues faced: Based on the reported information from the original paper, I was unable to interpret what the authors did with this analysis in sufficient detail to know how to reproduce the results. Had analysis code been provided this reproduction may have been more successful. Since there was no analysis code, I am unable to determine to what extent the failure to reproduce this KS test was due to a lack of understanding on my part, or a lack of clarity on the part of the authors - I believe both factors contributed to some extent.

ANOVA for reaction times (RTs)

```
##### CS ANOVA RTs data setup #####

# sept 2020 has missing data, exclude sept 2020 datapoints

cs_aovRT_data <- df_01 %>%
 dplyr::filter(
    time_as_number != 2
  ) # excludes sept 2020 datapoints

# original paper implies the same variables as the CS ANOVA, except the DV is
# now the RTs for inverted and upright. Therefore, need to pivot data to long
# format.

cs_aovRT_data <- cs_aovRT_data %>%
  tidyr::pivot_longer(
    cols = c("rt upright", "rt inverted"), # selects columns to be pivoted
    names_to = "rt_orientation",           # name of the long formatted column
    values_to = "rt"                       # values of the long formatted column
  ) # pivots data from wide to long format

# ANOVA requires rt, subjectID, group, gender, time, rt orientation

cs_aovRT_data <- cs_aovRT_data %>%
  select(
    subjectID, group, time, rt_orientation, rt
  ) # selects relevant columns for ANOVA
```

```
##### CS ANOVA RTs #####

cs_aovRT <- afex::aov_ez(
  id = "subjectID",             # independent variable
  dv = "rt",                    # dependent variable
  within = "rt_orientation",    # within-subject variables
```

```r
  between = c("group", "time"), # between-subject variables
  data = cs_aovRT_data,         # data
  anova_table = 'pes'           # we want partial eta squared ('pes'),
                                # is set to 'ges' by default
) # the ANOVA object

knitr::kable(
  nice(cs_aovRT),
  caption = "Cross-sectional ANOVA of RTs output (time points, orientation and group)."
  ) %>% # creates nice ANOVA table output
  kable_styling(
    latex_options = "HOLD_position"
    )
```

Table 4: Cross-sectional ANOVA of RTs output (time points, orientation and group).

Effect	df	MSE	F	pes	p.value
group	1, 1580	2944583.30	0.90	<.001	.344
time	4, 1580	2944583.30	2.26 +	.006	.061
group:time	4, 1580	2944583.30	0.39	<.001	.818
rt_orientation	1, 1580	891331.62	115.56 ***	.068	<.001
group:rt_orientation	1, 1580	891331.62	21.95 ***	.014	<.001
time:rt_orientation	4, 1580	891331.62	0.83	.002	.505
group:time:rt_orientation	4, 1580	891331.62	0.90	.002	.463

Original paper reported: "ANOVA demonstrated main effects of orientation (upright faster than inverted), F(1, 1570) = 113, p < .001, ηp^2 = .068. The interaction between orientation and group was also significant, with a reduced inversion effect for masked faces, F(1, 1570) = 22.4, p < .001, ηp^2 = .014, despite the lack of a main effect for group, F(1, 1570) <1. Importantly, there was no interaction between time point and group (mask status), F(1, 1570) < 1, suggesting that the mask effect remained constant across time points." (p. 1642).

Computational reproduction success: Results from this ANOVA were somewhat successfully reproduced (see table 4). Most discrepancies appear to be small, but degrees of freedom are wrong.

Issues faced: Whilst it was explicitly stated before the 'CS ANOVA', the exclusion of sept 2020 participants due to missing data from this 'ANOVA for RTs' was not stated. Only after running the ANOVA and encountering an error did I realise that the sept 2020 participants of course still had the missing data that required exclusion before running the ANOVA.

The reproduced F statistics do not exactly round to the original F statistics. Furthermore, there were discrepancies between the degrees of freedom for the reproduced and original results.

I'm not sure what has caused these discrepancies. They do not appear to have deviated much from the original results, but are different. Maybe the discrepancies are due to discrepancies between the R and JASP.

Bayesian ANOVA for RTs

```r
##### CS Bayesian ANOVA RTs #####

cs_bayes_RT <- df_01 %>%
  select(
```

```
    time, "rt upright", group # selects time and rt upright
  )

cs_bayes_RT$time <- as.factor(cs_bayes_RT$time)   # converts the time IV into factor
cs_bayes_RT$group <- as.factor(cs_bayes_RT$group) # converts the group IV into factor

anovaBF(
  formula = `rt upright` ~ time * group, # linear model formula
  data = cs_bayes_RT                     # the cs_bayes_up data
) %>%
  kable(
    caption = "Cross-sectional Bayesian ANOVA of time and RTs for upright faces."
) %>%
  kable_styling(
    latex_options = "HOLD_position"
    )
```

```
## Warning: data coerced from tibble to data frame
```

Table 5: Cross-sectional Bayesian ANOVA of time and RTs for upright faces.

	bf	error	time	code
time	1.613001	0.0000812	Thu Jan 26 09:49:14 2023	13667eb1553
group	3.576271	0.0000625	Thu Jan 26 09:49:14 2023	136374e8e15
time + group	1.189513	0.0104711	Thu Jan 26 09:49:14 2023	13670bdfee0
time + group + time:group	2.018394	0.0176236	Thu Jan 26 09:49:14 2023	1367ce87024

Original paper reported: "Bayesian ANOVA on the RT for upright faces provided support for the null hypothesis (BF10 = 0.084)." (p. 1642).

Computational reproduction success: Results were unsuccessfully reproduced.

Issues faced: Whilst I was able to run an ANOVA on this data, there were discrepancies. I don't think the original paper specified in enough detail what linear model was used in this Bayesian ANOVA. I'm not sure what the (bf10 = 0.084) statistic is actually referring to since only the outcome variable (RT) is mentioned.

Longitudinal Analysis

Repeated measures ANOVA (group, orientation, time)

```
##### LA RM ANOVA data setup #####

# original paper does not report any data exclusion

# lots of columns in df_02, select relevant columns first. There is also no time
# column in the dataset. Need to create a time column

la_aov_data <- df_02 %>%
  select(
    subjectID, upright, inverted, group, up2021, inv2021
  ) # selects relevant columns for ANOVA
```

```r
# original paper specifies that ANOVA involved time as a variable. There is no
# time variable in the original dataset. need to pivot then create time column.

# pivot upright and inverted CFMT scores for both 2021 and 2022.

la_aov_data <- la_aov_data %>%
  tidyr::pivot_longer(
    cols = c("up2021", "upright", "inv2021", "inverted"), # selects columns to
                                                          # be pivoted

    names_to = "orientation",                             # name of the long
                                                          # formatted column

    values_to = "CFMT"                                    # values of the long
                                                          # formatted column
  )

# create time column

la_aov_data <- la_aov_data %>%
  mutate(
    .,
    time = case_when(
      endsWith(orientation, "2021") ~ "2021",      # orientation = ends with "2021"
                                                   # becomes time = "2021"

      startsWith(orientation, "upright") ~ "2022", # orientation = "upright" becomes
                                                   # time = "2022"

      startsWith(orientation, "inverted") ~ "2022" # orientation = "inverted" becomes
                                                   # time = "2022"
    )
  ) # creates time column with value for each subject observation
```

```r
        ######################################
        ##### NOT FINISHED, error = TRUE #####
        ######################################

##### LA RM ANOVA #####

la_aov <- afex::aov_ez(
  id = "subjectID",                   # independent variable
  dv = "CFMT",                        # dependent variable
  within = c("orientation", "time"), # within-subject variables
  between = "group",               # between-subject variables
  data = la_aov_data,                 # data
  anova_table = 'pes'                 # we want partial eta squared ('pes'),
                                      # is set to 'ges' by default
) # the ANOVA object
```

```
## Error: Empty cells in within-subjects design (i.e., bad data structure).
## table(data[c("orientation", "time")])
## #            time
## # orientation X2021 X2022
```

```
## #    up2021     209     0
## #    upright      0   209
## #    inv2021    209     0
## #    inverted     0   209
```

```
knitr::kable(
  nice(la_aov),
  caption = "Longitudinal ANOVA for RTs output (group, orientation, time)."
  ) %>% # creates nice ANOVA table output
  kable_styling(
    latex_options = "HOLD_position"
    )
```

```
## Error in nice(la_aov): object 'la_aov' not found
```

Original paper reported: "A repeated measures ANOVA with group (masked, nonmasked), orientation (upright, inverted), and time point (2021, 2022) revealed a two-way interaction between time point and group, $F(1, 207) = 27.8$, $p < .001$, $\eta p^2 = .119$; for the nonmasked group, an improvement was observed for the second time point—upright: $F(1, 207) = 63.2$, $p < .001$; inverted: $F(1, 207) = 13.9$, $p < .001$. However, no improvement was found for the masked group (upright and inverted: $Fs < 1$). This interaction was qualified by a three-way interaction with orientation, $F(1, 207) = 3.82$, $p = .052$, $\eta p^2 = .018$, because the improvement for the nonmasked faces was slightly greater for upright faces" (p. 1643).

Computational reproduction success: Results were unsuccessfully reproduced. I was unable to determine how to structure the data for this ANOVA.

Issues faced: I found it challenging to know how to structure the data for this analysis so I was unable to reproduce the results.

ANOVA RTs

```
# original paper does not report any data exclusion

# lots of columns in df_02, select relevant columns first. There is also no time
# column in the dataset. Need to create a time column

la_aovRT_data <- df_02 %>%
  select(
    subjectID, RTupright, RTinverted, group, RTup2021, RTinv2021
  ) # selects relevant columns for ANOVA

# original paper specifies that ANOVA involved time as a variable. There is no
# time variable in the original dataset. need to pivot then create time column.

# pivot upright and inverted CFMT scores for both 2021 and 2022.

la_aovRT_data <- la_aovRT_data %>%
  tidyr::pivot_longer(
    cols = c(
      "RTup2021",
      "RTupright",
      "RTinv2021",
      "RTinverted"
      ), # selects columns to be pivoted
```

```r
    names_to = "orientation",
         # name of the long formatted column
    values_to = "RT"
         # values of the long formatted column

    ) # pivots data from wide to long format.

# create time column

la_aovRT_data <- la_aovRT_data %>%
  mutate(
    .,
    time = case_when(
      endsWith(orientation, "2021") ~ "2021",        # orientation = ends with "2021"
                                                      # becomes time = "2021"

      startsWith(orientation, "RTupright") ~ "2022", # orientation = "upright" becomes
                                                      # time = "2022"

      startsWith(orientation, "RTinverted") ~ "2022" # orientation = "inverted" becomes
                                                      # time = "2022"
    )
  ) # creates time column with value for each subject observation
```

```r
        ######################################
        ##### NOT FINISHED, error = TRUE #####
        ######################################

##### LA RM ANOVA RTs #####

la_aovRT <- afex::aov_ez(
  id = "subjectID",                 # independent variable
  dv = "RT",                        # dependent variable
  within = c("time", "orientation"), # within-subject variables
  between = c("group"),             # between-subject variables
  data = la_aovRT_data,             # data
  anova_table = 'pes'               # we want partial eta squared ('pes'),
                                    # is set to 'ges' by default
) # the ANOVA object
```

```
## Error: Empty cells in within-subjects design (i.e., bad data structure).
## table(data[c("time", "orientation")])
## #        orientation
## # time    RTup2021 RTupright RTinv2021 RTinverted
## #   X2021      209         0       209          0
## #   X2022        0       209         0        209
```

```r
knitr::kable(
  nice(la_aovRT),
  caption = "Longitudinal ANOVA for RTs output (group, orientation, time)."
  ) %>% # creates nice ANOVA table output
  kable_styling(
    latex_options = "HOLD_position"
    )
```

```
## Error in nice(la_aovRT): object 'la_aovRT' not found
```

Original paper reported: "The RT analysis revealed faster performance in January 2022 for both masked and nonmasked faces, $F(1, 207) = 31.6$, $p < .001$, $\eta p^2 = .13$. Notably, this effect was similar across the two categories (Time × Group interaction; $F < 1$), with no effect of group ($F < 1$)." (p. 1643).

Computational reproduction success: Results were unsuccessfully reproduced. I was unable to determine how to structure the data for this ANOVA.

Issues faced: I found it challenging to know how to structure the data for this analysis so I was unable to reproduce the results.

Correlation (Jan 2021, Jan 2022)

```
##### LA correlation (jan 2021, jan 2022) data setup #####

# original paper specifies upright masked faces and nonmasked faces found robust
# correlations

# need to pivot la_aov_data to wide in order to perform correlation for upright
# masked faces

la_cor <- la_aov_data %>%
  tidyr::pivot_wider(
    .,
    names_from = c("time", "orientation"), # both time and orientation need to be
                                           # pivoted
    values_from = "CFMT"
  ) %>% # pivots from long to wide
  select(
    -subjectID
  ) # subjectID not needed for correlation

# upright masked faces correlation

la_cor_up <- la_cor %>%
  dplyr::filter(
    group != "no mask"
  ) %>%
  select(
    -group, -"2021_inv2021", -"2022_inverted"
  )

# nonmasked faces correlation

la_cor_non <- la_cor %>%
  dplyr::filter(
    group != "mask"
  ) %>%
  select(
    -group, -"2021_inv2021", -"2022_inverted"
  )
```

```r
##### LA correlation (jan 2021, jan 2022) #####

# upright masked faces correlation

la_cor_mask_test <- la_cor_up %>%
  cor_test("2021_up2021", "2022_upright")

knitr::kable(
  la_cor_mask_test,
  caption = "upright masked faces correlation"
  ) %>%
  kable_styling(
    latex_options = "HOLD_position"
  )
```

Table 6: upright masked faces correlation

var1	var2	cor	statistic	p	conf.low	conf.high	method
2021_up2021	2022_upright	0.61	7.608285	0	0.4658866	0.7156991	Pearson

```r
# upright non-masked faces correlation

la_cor_non_test <- la_cor_non %>%
  cor_test("2021_up2021", "2022_upright")

knitr::kable(
  la_cor_non_test,
  caption = "upright non-masked faces correlation"
  ) %>%
  kable_styling(
    latex_options = "HOLD_position"
  )
```

Table 7: upright non-masked faces correlation

var1	var2	cor	statistic	p	conf.low	conf.high	method
2021_up2021	2022_upright	0.68	9.520088	0	0.5636217	0.7708852	Pearson

Original paper reported: "we invited participants who completed the experiment in January 2021 to retake the CFMT in January 2022… …we found robust correlations between the two sessions for both the upright masked faces ($r = .6$, $p <.001$) and nonmasked faces ($r = .68$, $p < .001$)" (p. 1643).

Computational reproduction success: Results for these correlations were almost entirely successfully reproduced (see tables). Discrepancies appear to be due to rounding errors.

Issues faced: The description of the correlation provided was somewhat ambiguous to me initially. I first interpreted the description as suggesting the first correlation involved the upright masked faces condition and the second correlation involved both upright and inverted nonmasked conditions. However, the description actually says that both correlations involved the upright faces, the first involving masked faces and the second involving nonmasked faces. Whilst this wording did not ultimately prevent reproduction of the results, it introduced unnecessary confusion. I#m also unsure to what extent the description was unclear or to what extent I was unable to interpret the description.

The EMF Questionnaire

EMF ANOVA

```
##### EMF ANOVA #####

# Subjective ~ WorkLoc
# select subjectID, WorkLoc and subjective

EMF_anova_df_1 <- df_02 %>%
  select(
    subjectID, WorkLoc, subjective
  )

afex::aov_ez(
  id = "subjectID",
  dv = "subjective",
  between = "WorkLoc",
  data = EMF_anova_df_1,
  anova_table = "pes"
)
```

```
## Converting to factor: WorkLoc

## Contrasts set to contr.sum for the following variables: WorkLoc

## Anova Table (Type 3 tests)
##
## Response: subjective
##    Effect     df  MSE    F  pes p.value
## 1 WorkLoc 2, 206 9.12 1.13 .011    .324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
# Prop ~ WorkLoc
# select subjectID, WorkLoc and Prop

EMF_anova_df_2 <- df_02 %>%
  select(
    subjectID, WorkLoc, Prop
  )

afex::aov_ez(
  id = "subjectID",
  dv = "Prop",
  between = "WorkLoc",
  data = EMF_anova_df_2,
  anova_table = "pes"
)
```

```
## Converting to factor: WorkLoc
## Contrasts set to contr.sum for the following variables: WorkLoc

## Anova Table (Type 3 tests)
```

```
##
## Response: Prop
##    Effect     df    MSE      F  pes p.value
## 1 WorkLoc 2, 206 642.07 3.76 * .035    .025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

Original paper reported: "participants who work from their office reported greater experience with masked faces (Items 5 and 6), F(2, 701) = 21, p < .001, ηp^2 = .057, and more encounters with masked individuals (Item 9), F(2, 701) = 4.89, p < .01, ηp^2 = .01." (p. 1643).

Computational reproduction success: Results were unsuccessfully reproduced.

Issues faced: Whilst I was able to run an ANOVA on this data, there were discrepancies. The first result says "reported greater experience with masked faces" but refers to items from the subjective scale (items 5 and 6) of the EMF rather than the experience scale which the description implies. I was confused as to which scale they were actually referring to here. Furthermore, discrepancies in the degrees of freedom imply that different data was used between the original analysis and the reproduced analysis. I am unsure how to approach this analysis beyond what was attempted.

Repeated measures ANOVA (group, orientation, work location)

```
##### EMF RM ANOVA #####

EMF_rm <- df_02 %>%
  pivot_longer(
    cols = c("upright", "inverted"),
    names_to = "orientation",
    values_to = "CFMT"
  )

afex::aov_ez(
  id = "subjectID",
  dv = "CFMT",
  between = c("group", "WorkLoc"),
  within = "orientation",
  data = EMF_rm,
  anova_table = "pes"
)
```

```
## Converting to factor: group, WorkLoc

## Contrasts set to contr.sum for the following variables: group, WorkLoc

## Anova Table (Type 3 tests)
##
## Response: CFMT
##                       Effect     df    MSE          F  pes p.value
## 1                      group 1, 203 103.11  60.07 *** .228   <.001
## 2                    WorkLoc 2, 203 103.11       2.24 .022    .110
## 3              group:WorkLoc 2, 203 103.11       0.58 .006    .562
## 4                orientation 1, 203  31.69 494.32 *** .709   <.001
## 5          group:orientation 1, 203  31.69  47.49 *** .190   <.001
## 6        WorkLoc:orientation 2, 203  31.69       1.03 .010    .360
## 7 group:WorkLoc:orientation 2, 203  31.69       0.96 .009    .385
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

Original paper reported: "repeated measures ANOVA with group (masked, nonmasked), orientation (upright, inverted), and work location (home, office, unemployed). Whereas there was a robust main effect of masks, $F(1, 686) = 143$, $p < .001$, $\eta p^2 = .17$, there was no interaction with work location, $F(2, 686) = 2.03$, $p = .132$, $\eta p^2 = .005$" (p. 1643-1644).

Computational reproduction success: Results were unsuccessfully reproduced.

Issues faced: Whilst I was able to run the ANOVA with the data, there were discrepancies. Partial eta squared values appear to be very similar but F statistics, degrees of freedom and P-values are wrong. I am unsure how to approach this analysis beyond what was attempted.

Bayesian ANOVA (masked upright)

```
##### EMF Bayesian ANOVA #####

EMF_bayes <- df_02 %>%
  filter(
    group == "mask"
  ) %>%
  select(
    group, upright, WorkLoc
  )

EMF_bayes$WorkLoc <- as.factor(EMF_bayes$WorkLoc) # converts the time IV into factor

anovaBF(
  formula = upright ~ WorkLoc, # linear model formula
  data = EMF_bayes             # the cs_bayes_up data
) %>%
  kable(
    caption = "EMF Bayesian ANOVA of work location for upright faces."
) %>%
  kable_styling(
    latex_options = "HOLD_position"
    )
```

```
## Warning: data coerced from tibble to data frame
```

Table 8: EMF Bayesian ANOVA of work location for upright faces.

	bf	error	time	code
WorkLoc	0.3267706	0.0003074	Thu Jan 26 09:49:15 2023	1367027eb76

Original paper reported: "Bayesian ANOVA on the accuracy score for masked upright faces further established the lack of work location effect (BF10 = 0.119; the null hypothesis is 7.4 times more probable than the alternative hypothesis)" (p. 1644).

Computational reproduction success: Results were unsuccessfully reproduced.

Issues faced: Whilst I was able to run the bayesian ANOVA on the data, there were discrepancies. The description seemed to suggest the Bayesian ANOVA only involved upright accuracy scores as the outcome variable and work location as the predictor. I was unsure how to approach this analysis beyond what was attempted.

Correlation (Masked and Masked-Unmasked)

Note: The original paper specifies that correlations were only conducted with participants from the masked condition. However, they also claim that the results from this correlational analysis was similar to the results when both masked and unmasked conditions were taken together. Therefore, this computational reproduction will look at: 1) masked on its own 2) masked and unmasked together.

```
##### EMF correlation masked only #####

EMF_cor_mask <- df_02 %>%
  filter(
    group == "mask"
  ) %>%
  select(
    upright, inverted, RTupright, RTinverted, experience, regulation, subjective, Prop, PI20
  )

rcorr(
  as.matrix(EMF_cor_mask)
  ) %>%
  broom::tidy() %>%
  print(n = Inf)
```

```
## # A tibble: 36 x 5
##    column1    column2    estimate     n  p.value
##    <chr>      <chr>         <dbl> <int>    <dbl>
##  1 inverted   upright     0.553     102 1.62e- 9
##  2 RTupright  upright     0.0580    102 5.63e- 1
##  3 RTupright  inverted    0.0728    102 4.67e- 1
##  4 RTinverted upright     0.197     102 4.76e- 2
##  5 RTinverted inverted    0.321     102 9.85e- 4
##  6 RTinverted RTupright   0.402     102 2.87e- 5
##  7 experience upright    -0.0447    102 6.56e- 1
##  8 experience inverted    0.0131    102 8.96e- 1
##  9 experience RTupright  -0.00984   102 9.22e- 1
## 10 experience RTinverted -0.0793    102 4.28e- 1
## 11 regulation upright     0.162     102 1.04e- 1
## 12 regulation inverted    0.154     102 1.21e- 1
## 13 regulation RTupright   0.0260    102 7.96e- 1
## 14 regulation RTinverted -0.00792   102 9.37e- 1
## 15 regulation experience  0.395     102 4.07e- 5
## 16 subjective upright    -0.271     102 5.93e- 3
## 17 subjective inverted   -0.110     102 2.71e- 1
## 18 subjective RTupright   0.126     102 2.05e- 1
## 19 subjective RTinverted -0.0120    102 9.05e- 1
## 20 subjective experience -0.161     102 1.06e- 1
## 21 subjective regulation -0.179     102 7.21e- 2
## 22 Prop       upright     0.0367    102 7.14e- 1
## 23 Prop       inverted   -0.0392    102 6.95e- 1
```

```
## 24 Prop       RTupright   0.000654   102 9.95e- 1
## 25 Prop       RTinverted -0.122      102 2.22e- 1
## 26 Prop       experience  0.633      102 9.37e-13
## 27 Prop       regulation  0.488      102 1.99e- 7
## 28 Prop       subjective -0.193      102 5.19e- 2
## 29 PI20       upright    -0.350      102 3.16e- 4
## 30 PI20       inverted   -0.137      102 1.69e- 1
## 31 PI20       RTupright   0.182      102 6.69e- 2
## 32 PI20       RTinverted -0.0492     102 6.23e- 1
## 33 PI20       experience -0.117      102 2.44e- 1
## 34 PI20       regulation -0.137      102 1.71e- 1
## 35 PI20       subjective  0.589      102 7.26e-11
## 36 PI20       Prop       -0.00624    102 9.50e- 1
```

```
##### EMF correlation for both masked and unmasked #####

EMF_cor_both <- df_02 %>%
  select(
    upright, inverted, RTupright, RTinverted, experience, regulation, subjective, Prop, PI20
  )

rcorr(
  as.matrix(EMF_cor_both)
  ) %>%
  broom::tidy() %>%
  print(n = Inf)
```

```
## # A tibble: 36 x 5
##    column1    column2    estimate     n  p.value
##    <chr>      <chr>         <dbl> <int>    <dbl>
##  1 inverted   upright     0.584     209 0
##  2 RTupright  upright    -0.0405    209 5.60e- 1
##  3 RTupright  inverted    0.0453    209 5.14e- 1
##  4 RTinverted upright     0.123     209 7.57e- 2
##  5 RTinverted inverted    0.177     209 1.02e- 2
##  6 RTinverted RTupright   0.437     209 3.74e-11
##  7 experience upright    -0.00341   209 9.61e- 1
##  8 experience inverted    0.0211    209 7.62e- 1
##  9 experience RTupright  -0.106     209 1.25e- 1
## 10 experience RTinverted -0.114     209 9.97e- 2
## 11 regulation upright     0.106     209 1.26e- 1
## 12 regulation inverted    0.145     209 3.59e- 2
## 13 regulation RTupright  -0.0711    209 3.06e- 1
## 14 regulation RTinverted -0.0571    209 4.12e- 1
## 15 regulation experience  0.533     209 0
## 16 subjective upright    -0.260     209 1.45e- 4
## 17 subjective inverted   -0.152     209 2.83e- 2
## 18 subjective RTupright   0.156     209 2.44e- 2
## 19 subjective RTinverted -0.0574    209 4.09e- 1
## 20 subjective experience -0.121     209 8.03e- 2
## 21 subjective regulation -0.0755    209 2.78e- 1
## 22 Prop       upright     0.0155    209 8.24e- 1
## 23 Prop       inverted   -0.0267    209 7.01e- 1
## 24 Prop       RTupright  -0.0630    209 3.65e- 1
## 25 Prop       RTinverted -0.0982    209 1.57e- 1
```

```
## 26 Prop       experience  0.675     209 0
## 27 Prop       regulation  0.525     209 4.44e-16
## 28 Prop       subjective -0.0743    209 2.85e- 1
## 29 PI20       upright    -0.328     209 1.27e- 6
## 30 PI20       inverted   -0.228     209 8.84e- 4
## 31 PI20       RTupright   0.144     209 3.72e- 2
## 32 PI20       RTinverted -0.0825    209 2.35e- 1
## 33 PI20       experience -0.0949    209 1.72e- 1
## 34 PI20       regulation -0.114     209 1.01e- 1
## 35 PI20       subjective  0.530     209 0
## 36 PI20       Prop       -0.00107   209 9.88e- 1
```

Original paper reported: "The next analysis was conducted only on the participants who completed the mask condition (similar results were obtained when both groups were included). We found a robust correlation between the Experience scale (Items 1, 2, 7, and 8) and the continuous evaluation of experience with masked faces (Item 9; $r = .494$, $p < .001$), between the Experience and Regulation scales ($r = .51$, $p < .001$), and between the Regulation scale and the continuous evaluation of experience with masked faces ($r = .52$, $p < .001$). As expected, we also found strong correlations between performance with upright and inverted faces for both accuracy and RT. Importantly, there were no correlations between face-perception abilities and experience with masked faces. This was true for accuracy scores and RTs for upright and inverted faces (all $ps > .1$; Table 3). Participants from January 2022 also completed the PI20, a validated questionnaire of face-perception abilities (Shah, Sowden, et al., 2015). The PI20 score showed a significant correlation with the subjective scale of the EMF questionnaire ($r = .619$, $p < .001$). (p. 1644).

Computational reproduction success: Whilst the significance, or non-significance, of each reported result was reproduced qualitatively, neither the exact p-values nor the correlational coefficients were reproduced.

Issues faced: Whilst the paper says they conducted the analysis on the masked condition only, they also claim correlations were similar between just the masked condition and both masked and unmasked combined. The authors do not provide any correlation analysis output for the masked and unmasked combined correlations, so there is ambiguity in what is meant by 'similar' here and how the authors made that assessment.

The general pattern of significance (or non-significance) described in the paper was reproduced entirely, however very few of the reproduced p-values matched the reported p-values, nor were any of the correlational coefficients reproduced. I am unsure why this is the case.

Relevant values, like sample size, were not included in the original results which may account for why discrepancies emerged.

CFMT: within-subject design

Repeated measures ANOVA (mask status as within-subjects variable)

```
##### CFMT RM ANOVA #####

# RM ANOVA
CFMT_df <- df_03 %>%
  select(
    subjectID, no_mask, mask
  ) %>%
  pivot_longer(
    cols = c("no_mask", "mask"),
    names_to = "mask_status",
```

```r
    values_to = "CFMT"
  )

CFMT_anova <- afex::aov_ez(
  id = "subjectID",
  dv = "CFMT",
  within = "mask_status",
  data = CFMT_df,
  anova_table = "pes"
)

kable(
  nice(CFMT_anova)
  ) %>%
  kable_styling(
    latex_options = "HOLD_position"
  )
```

Effect	df	MSE	F	pes	p.value
mask_status	1, 199	31.17	307.45 ***	.607	<.001

```r
# CFMT accuracy means

mean(df_03$mask)
```

```
## [1] 44.23
```

```r
# 44.23

mean(df_03$no_mask)
```

```
## [1] 54.02
```

```r
# 54.02
```

Original paper reported: "A repeated measures ANOVA with mask status as a within-subjects variable revealed a robust main effect, F(1, 299) = 499, p < .001, $\eta p^2 = .626$, with greater accuracy for nonmasked faces (CFMT nonmasked: 54.4; CFMT masked: 44.81; mask effect = ~18%)." (p. 1645).

Computational paper reported: Results were somewhat reproduced successfully. The general pattern was reproduced but specific values (F-statistics, degrees of freedom, p-values etc.) were not reproduced exactly. The reported means averages were also not reproduced.

Issues faced: I was able to reproduce the same overall pattern of results, but inspecting the ANOVA output revealed that specific test statistics were not reproduced. I do not know what caused the discrepancies or why despite these discrepencies the overall pattern of results was reproduced.

Mean CFMT scores for masked and nonmasked groups were almost identical, but not exactly. The mean averages should really be identical and the presence of discrepancies suggests (slight) differences between the data used in the original paper and that provided on OSF.

ANOVA RTs

```r
##### CFMT ANOVA RTs #####

# CFMT ANOVA RTs

CFMT_df_2 <- df_03 %>%
  select(
    subjectID, no_mask_RT, mask_RT
  ) %>%
  pivot_longer(
    cols = c("no_mask_RT", "mask_RT"),
    names_to = "mask_status",
    values_to = "CFMT"
  )

CFMT_anova_RT <- afex::aov_ez(
  id = "subjectID",
  dv = "CFMT",
  within = "mask_status",
  data = CFMT_df_2,
  anova_table = "pes"
)

kable(
  nice(CFMT_anova_RT)
  ) %>%
  kable_styling(
    latex_options = "HOLD_position"
  )
```

Effect	df	MSE	F	pes	p.value
mask_status	1, 199	8820290.81	13.62 ***	.064	<.001

```r
# CFMT RTs Means

mean(df_03$mask_RT)
```

```
## [1] 4962.839
```

```r
# 4962.839

mean(df_03$no_mask_RT)
```

```
## [1] 3866.661
```

```r
# 3866.661
```

Original paper reported: "RT analysis corroborated this finding with longer RTs for masked faces, F(1, 299) = 19.2, p < .001, ηp^2 = .06 (CFMT non-masked: 3,631 ms; CFMT masked: 4,539 ms)." (p. 1645).

Computational reproduction success: Results were somewhat reproduced successfully. The general pattern was reproduced by specific values (F-statistics, degrees of freedom, p-values etc.) were not reproduced exactly. The reported mean averages were also not reproduced.

Issues faced: I was able to reproduce the same overall pattern of results, but inspecting the ANOVA output revealed that specific test statistics were not reproduced. I do not know what caused the discrepancies or why despite these discrepencies the overall pattern of results was reproduced.

Discrepancies between the reported mean averages and the reproduced means suggest different data was used between the original and reproduced analyses. Authors should have been more explicit about which data was used.

Correlation (mask effect [accuracy and RTs], longitudinal EMF scores)

```
##### CFMT correlation #####
```

Original paper reported: "we examined the relationship between the mask effect (in terms of accuracy and RT) and the scores obtained from the EMF questionnaire for the longitudinal data (Table 4). As expected, we found a robust correlation between the Experience scale (Items 1, 2, 7, and 8) and the continuous evaluation of experience with masked faces (Item 9; $r = .64$, $p < .001$). We also found strong correlations between the Experience and Regulation (Items 3 and 4) scales ($r = .41$, $p < .001$) and between the Regulation scale and the continuous evaluation of experience with masked faces ($r = .46$, $p < .001$).Importantly, we did not find any correlations between the mask effect (in terms of either accuracy or RT) and experience with masked faces (all $ps > .1$)." (p. 1645-1646).

Computational reproduction success: Results were unsuccessfully reproduced. The data used in this analysis was difficult to interpret so I was not able to perform a reproduction of the analysis.

Issues faced: From the description in the original paper, I was unable to determine what data these results referred to. As such I could not work out how to approach reproducing the results.

ANOVA occupations (health & retail, students & unemployed)

```
##### CFMT ANOVA occupations #####
```

Original paper reported: "we collected information about the occupational profile of our participants. We then focused our analysis on occupations that are more likely to have an increased amount (health and retail workers; $n = 46$) or reduced amount (students and unemployed; $n = 124$) of masked-face experiences. First, as expected, we found that health and retail workers had more experience with masked faces—Experience scale: $F(3, 166) =8.33$, $p < .001$, $\eta p^2 = .13$; percentage-of-masked-faces scale: $F(3, 166) = 8.54$, $p < .001$, $\eta p^2 = .13$." (p. 1646).

Computational reproduction success: Results were unsuccessfully reproduced. This was attributed to a lack of data.

Issues faced: I'm unsure where in the data set information about employment or education can be found beyond work location (unemployed, office, home). As such I do not know if there is data available to reproduce these results.

ANOVA (workplace, mask status)

```
##### CFMT ANOVA workplace & mask status #####
```

Original paper reported: "Critically, an additional ANOVA with workplace and mask status showed a robust main effect for mask status, $F(1, 166) =22.3$, $p < .001$, $\eta p^2 = .119$, but no main effect for workplace ($F < 1$) nor an interaction between these factors ($F < 1$)" (p. 1646).

Computational reproduction success: Results were unsuccessfully reproduced. This was attributed to a lack of data.

Issues faced: I'm unsure where in the data set information about employment or education can be found beyond work location (unemployed, office, home). As such I do not know if there is data available to reproduce these results.

GFMT: masked faces

Correlation (face-perception performance [d' & RTs], EMF scores)

```
##### GFMT correlation face=perception and EMF score #####

GFMT_cor <- df_04 %>%
  select(
    -subjectID, -age, -gender, -WorkLocation
  )

rcorr(
  as.matrix(GFMT_cor)
  ) %>%
  broom::tidy() %>%
  print(n = Inf)
```

```
## # A tibble: 78 x 5
##    column1      column2      estimate     n  p.value
##    <chr>        <chr>           <dbl> <int>    <dbl>
##  1 Different    Same         -0.291     284 5.74e- 7
##  2 Same_RT      Same          0.191     284 1.20e- 3
##  3 Same_RT      Different     0.0303    284 6.11e- 1
##  4 Different_RT Same         -0.0790    284 1.85e- 1
##  5 Different_RT Different     0.278     284 1.92e- 6
##  6 Different_RT Same_RT       0.719     284 0
##  7 q1           Same         -0.0972    284 1.02e- 1
##  8 q1           Different     0.123     284 3.85e- 2
##  9 q1           Same_RT      -0.0555    284 3.51e- 1
## 10 q1           Different_RT -0.00913   284 8.78e- 1
## 11 q2           Same         -0.0667    284 2.63e- 1
## 12 q2           Different     0.00166   284 9.78e- 1
## 13 q2           Same_RT      -0.0139    284 8.16e- 1
## 14 q2           Different_RT  0.0390    284 5.13e- 1
## 15 q2           q1            0.564     284 0
## 16 q3           Same         -0.0120    284 8.40e- 1
## 17 q3           Different     0.0390    284 5.13e- 1
## 18 q3           Same_RT       0.233     284 7.14e- 5
## 19 q3           Different_RT  0.194     284 9.94e- 4
## 20 q3           q1            0.202     284 5.98e- 4
## 21 q3           q2            0.332     284 1.03e- 8
## 22 q4           Same          0.0201    284 7.36e- 1
## 23 q4           Different    -0.0499    284 4.02e- 1
## 24 q4           Same_RT       0.128     284 3.13e- 2
## 25 q4           Different_RT  0.114     284 5.55e- 2
## 26 q4           q1            0.175     284 3.04e- 3
## 27 q4           q2            0.214     284 2.90e- 4
```

```
## 28 q4          q3            0.615     284 0
## 29 q5          Same          0.0117    284 8.44e- 1
## 30 q5          Different    -0.115     284 5.31e- 2
## 31 q5          Same_RT       0.0296    284 6.19e- 1
## 32 q5          Different_RT -0.0779    284 1.91e- 1
## 33 q5          q1           -0.0892    284 1.34e- 1
## 34 q5          q2            0.0122    284 8.37e- 1
## 35 q5          q3            0.00926   284 8.77e- 1
## 36 q5          q4            0.0247    284 6.78e- 1
## 37 q6          Same          0.0352    284 5.55e- 1
## 38 q6          Different    -0.0710    284 2.33e- 1
## 39 q6          Same_RT       0.0546    284 3.59e- 1
## 40 q6          Different_RT -0.0406    284 4.96e- 1
## 41 q6          q1           -0.0560    284 3.47e- 1
## 42 q6          q2           -0.0500    284 4.01e- 1
## 43 q6          q3            0.0331    284 5.79e- 1
## 44 q6          q4            0.0642    284 2.81e- 1
## 45 q6          q5            0.644     284 0
## 46 q7          Same         -0.0313    284 6.00e- 1
## 47 q7          Different     0.00138   284 9.81e- 1
## 48 q7          Same_RT       0.103     284 8.35e- 2
## 49 q7          Different_RT  0.0814    284 1.71e- 1
## 50 q7          q1            0.277     284 2.12e- 6
## 51 q7          q2            0.398     284 3.24e-12
## 52 q7          q3            0.653     284 0
## 53 q7          q4            0.467     284 0
## 54 q7          q5            0.0344    284 5.64e- 1
## 55 q7          q6            0.0634    284 2.87e- 1
## 56 q8          Same         -0.0702    284 2.38e- 1
## 57 q8          Different     0.0117    284 8.44e- 1
## 58 q8          Same_RT       0.0610    284 3.05e- 1
## 59 q8          Different_RT  0.0873    284 1.42e- 1
## 60 q8          q1            0.301     284 2.32e- 7
## 61 q8          q2            0.483     284 0
## 62 q8          q3            0.382     284 2.65e-11
## 63 q8          q4            0.263     284 6.80e- 6
## 64 q8          q5            0.0305    284 6.09e- 1
## 65 q8          q6           -0.0240    284 6.87e- 1
## 66 q8          q7            0.459     284 4.44e-16
## 67 q9          Same         -0.0750    284 2.08e- 1
## 68 q9          Different     0.0506    284 3.95e- 1
## 69 q9          Same_RT       0.154     284 9.38e- 3
## 70 q9          Different_RT  0.149     284 1.20e- 2
## 71 q9          q1            0.265     284 6.23e- 6
## 72 q9          q2            0.382     284 2.70e-11
## 73 q9          q3            0.569     284 0
## 74 q9          q4            0.400     284 2.37e-12
## 75 q9          q5           -0.0259    284 6.63e- 1
## 76 q9          q6           -0.0100    284 8.66e- 1
## 77 q9          q7            0.779     284 0
## 78 q9          q8            0.497     284 0
```

Original paper reported: "we examined the relationship between face-perception performance (d and RT) and the scores obtained from the EMF questionnaire (Table 5). Similar to the results observed for the

CFMT (see above), we found a robust correlation between the Experience scale (Items 1, 2, 7, and 8) and the continuous evaluation of experience with masked faces (Item 9; $r = .65$, $p < .001$), between the Experience and Regulation scales ($r = .51$, $p < .001$), and between the Regulation scale and the continuous evaluation of experience with masked faces ($r = .54$, $p < .001$). Importantly, once again, we did not find any correlations between face-perception abilities and experience with masked faces (all $ps > .1$). There were weak, yet significant correlations between RT and the Regulation scale ($r = .2$) as well as for RT and the continuous evaluation of experience with masked faces ($r = .16$); participants who reported greater experience with masked faces had slower RTs" (p. 1646).

Computational reproduction success: Results were partially reproduced. Due to how the dataset is structured, I was not able to reproduce the results for composite measures but did reproduce the general pattern of results for individual scale items.

Issues faced: The authors report correlations between composite measures (e.g. experience scale (items 1, 2, 7 and 8) and continuous evaluation of experience with masked faces (item 9)) but the dataset only includes individual scale items. Whilst correlations could be found between individual scale items, due to there being no composite measures in the dataset, I was unable to reproduce the results as reported when composite measures were involved.

Conclusion

Overall, this study reported 18 different analyses. Of these 18, only 3 were successfully reproduced entirely. An overarching issue with reproducing analyses for this study was the lack of analysis code. Authors claimed to have provided this code on OSF, but this was not the case. As such, reproducing the analyses required me to try to intepret how the analysis was conducted using what was reported in the paper. Had there been analysis code provided, it would likely have been easier to attempt reproductions of these results. Furthermore, some analyses in this study (i.e. the Bayesian ANOVAs) were conducted using JASP and information regarding the analysis decisions made in JASP was also not provided. These overarching issues meant analysis decision had to be inferred from the reported results, which led to further issues.

One such issue was the lack of detail about the data used in each analysis. Both the specific datasets used and how the data was structured for each analysis was unclear with almost all analyses. How to organise the data such that it was ready to be analysed had to be inferred for all analyses, often leading to discrepancies between the reported results and the reproduced results.

Another issue faced was a lack of clarity about the linear models used in the ANOVA analyses. It was often the case that the descriptions provided in the results section did not provide sufficient detail to understand what models the ANOVAs were actually being conducted on. As such, this often lead to discrepancies between the reported results and reproduced results. For the KS test, the description provided was not clear enough to make reproduction possible.

A further issue faced was a lack of clarity in the reporting of test statistics. There were several cases where reported statstics were combined together and summarised as being above or below some threshold value (e.g. all ps. $> .1$). Whilst in some cases these could be corroborated by reproduced analyses, it was hard to judge whether the analyses were actually reproduced or if they were instead within the same ballpark but were actually incorrect. Furthermore, some relevant test statistics were neglected in the original results, such as sample sizes for correlational analyses, which further added to difficulties in reproducing results.

There were 2 analyses which referred to data which was not present in any of the provided datasets and as such could not be reproduced.

In general, this paper often failed to provide sufficient detail to make reproduction of analyses possible. Beyond the lack of analysis code being provided, the reporting of results often ommitted necessary details for computational reproduction. Authors should have:

- Provided the original analysis code on OSF.

- Been explicit about the models and variables being used in ANOVAs and Bayesian ANOVAs.
- Explicitly referred to the specific data used in each analysis and how the data was structured for the analysis, rather than just providing raw datasets.
- Provided more detailed test statistics to make it easier to determine whether results were actually reproduced or not.

References

Freud, Erez, Daniela Di Giammarino, Andreja Stajduhar, R. Shayna Rosenbaum, Galia Avidan, and Tzvi Ganel. 2022. "Recognition of Masked Faces in the Era of the Pandemic: No Improvement Despite Extensive Natural Exposure." *Psychological Science* 33 (10): 1635–50. https://doi.org/10.1177/09567976221105459.