Recognition of masked faces in the era of the pandemic: No improvement, despite extensive, natural exposure

Preprint · June 2021

DOI: 10.31234/osf.io/x3gzq

5 authors, including:

Erez Freud
York University

Andreja Stajduhar
University of Toronto

Galia Avidan
Ben-Gurion University of the Negev

Tzvi Ganel
Ben-Gurion University of the Negev

Recognition of masked faces in the era of the pandemic: No improvement, despite extensive, natural exposure

Erez Freud[1], Andreja Stajduhar[1], Daniela Di Giammarino[1] R. Shayna Rosenbaum[1,2], Galia Avidan[3,4] and Tzvi Ganel[3]

[1]Department of Psychology and the Centre for Vision Research, York University, Toronto, Canada

[2]Rotman Research Institute, Baycrest Health Sciences, Toronto, Canada

[3]Psychology Department, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel

[4]Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel

Corresponding author: Erez Freud, efreud@yorku.ca

Keywords: Face Perception, COVID-19, Inversion Effect, learning and training,

Author contributions: EF, TG, RSR and GA developed the study concept and contributed to study design. AS collected the data. EF analyzed the data. EF drafted the first version of the manuscript. TG, RSR and GA edited the manuscript.

Acknowledgements: This study was supported by the Vision Science to Applications (VISTA) program funded by the Canada First Research Excellence Fund (CFREF, 2016–2023) (EF, RSR) by the Natural Sciences and Engineering Research Council (EF).

Abstract (149 words)

Face masks became prevalent across the globe to minimize the effects of the COVID-19 pandemic. Previous research highlighted their negative qualitative and quantitative impact on face recognition. An outstanding question is whether these effects would attenuate following extended natural exposure to masked faces. This question also pertains to potential effects of training on face recognition in natural settings. We utilized the Cambridge Face Memory Test (CFMT) in a cross-sectional study (total N=1,437), at five different time points over a 16-month period. The results showed persistent deficits in recognizing masked faces across time points. This was followed by a qualitative change implicated in a reduced inversion effect for masked faces. Additional experiments verified that the amount of individual experience with masked faces was not correlated with the mask effect. These findings provide compelling evidence that the mature face processing system in humans can be rigid even with prolonged, real-life exposure.

Statement of Relevance

Face recognition is one of the most skilled abilities in human perception. But what happens in the face of a sudden change, when faces appear dramatically different? This occurred during the COVID-19 pandemic: faces were frequently occluded by masks, which led to deficits in our ability to recognize people. However, the ongoing nature of the pandemic provided an unprecedented opportunity to examine the plasticity of the mature face processing system; in other words, did recognition of masked faces improve? Here, we measured the effect of masks on face recognition at five different time points over a 16-month period in a cross-sectional study (total-N=1,437). We found robust evidence for no improvement in recognizing masked faces with masks, even when taking into account individual's relative experience with masked faces. These findings provide compelling support for the idea that the mature face processing system is rigid in nature, even with prolonged, real-life exposure to altered faces.

Introduction

Faces are among the most informative visual stimuli in human perception, as they provide vital information for social interactions. Humans can extract significant information from brief exposure to a person's face, including their identity, gender, emotion, age, and race. While it is clear that face perception abilities rapidly develop from early childhood to adolescence (Pascalis et al., 2011), it is uncertain whether these abilities can improve later in life, particularly under situations in which visual input is sub-optimal (Pascalis et al., 2020; White, Kemp, Jenkins, Matheson, et al., 2014; Yovel et al., 2012). Here, we investigated this issue in a large-scale naturalistic study that tracked the effects of non-medical masks on face perception over a period of 16 months during the COVID-19 pandemic.

The COVID-19 pandemic introduced an unprecedented reality in which mask wearing became prevalent, and often mandatory, around the globe (Figure 1). Previous research demonstrated that the occlusion of the lower part of the face hinders different aspects of face processing, including recognition of familiar faces (Carragher & Hancock, 2020), unfamiliar faces (Dhamecha et al., 2014; Freud et al., 2020; Gosselin & Schyns, 2001; Kret & De Gelder, 2012; Stajduhar et al., 2021), and emotional expressions (Grundmann et al., 2021). Notably, previous studies were conducted before, or soon after, face masks became widely prevalent (i.e., first and second quarter of 2020), when participants did not have substantial experience with masked faces. Hence, an outstanding question is whether such extensive and prolonged experience with masked faces would change the way masked faces are perceived. Alongside its relevance and timeliness, this research also pertains to an outstanding theoretical question regarding the plasticity of the matured human face-processing system.

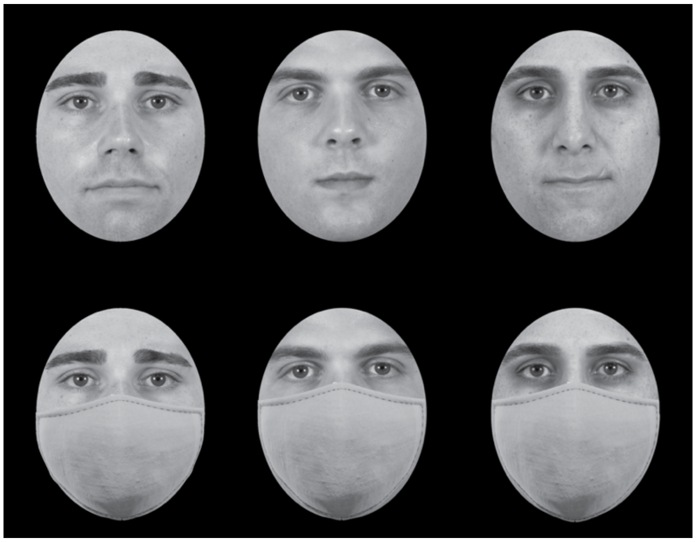

Figure 1: Examples of faces with and without masks similar to the ones used in the experiment. Faces are reproduced with permission from the Chicago Face Database (Ma et al., 2015)

Sensitivity to faces emerges already in newborns who exhibit a preference to face-like stimuli (Johnson & Morton, 1991) and to familiar faces over unfamiliar faces (Pascalis et al., 1995). Following this early sensitivity, the face processing system is subject to a prolonged developmental trajectory that ends after puberty (Chung & Thomson, 1995; Pascalis et al., 2011). During development, experience with faces plays an important and necessary role in refining face processing abilities. For example, the Other Race Effect (ORE; better recognition of own-race faces compared to faces of a different race) is observed in three-month-old infants after they gain some experience with faces of their own race (Kelly et al., 2005, 2009). Real-life experience can mitigate the ORE such that children who were adopted and raised by families from a different culture exhibit a reduced or even reversed ORE (De Heering et al., 2010; Sangrigoli et al., 2005). Finally, there is evidence that monkeys raised without exposure to faces do not develop face-selective patches in their brains nor preferential looking behaviors for faces, demonstrating the necessity of experience with faces to the formation of a functioning face processing system (Arcaro et al., 2017). This conclusion is corroborated by studies that showed altered face processing behaviors in adults who suffered from congenital cataracts that prevented visual input in the initial weeks after birth (Maurer, 2017).

The combination of a protracted developmental trajectory and sensitivity to experience supports the notion that training or exposure can sculpt face perception abilities during childhood (Pascalis et al., 2020). Accordingly, it was found that nine-month-old human infants developed better recognition abilities of monkey faces following individuation of these faces (Pascalis et al., 2005; Scott & Monesson, 2009). Additionally, reduced ORE was found among children who were exposed to faces from other races as long as the extensive exposure occurred before the age of 12 (McKone et al., 2019). Finally, training programs with neurotypical children (Bate et al., 2020) and children who suffer from prosopagnosia (Schmalzl et al., 2008) show consistent improvement in face perception abilities, demonstrating the flexibility of the developing face processing system. Notably, however, these latter effects are rather small and the degree of generalization to other situations and tasks is unclear.

In contrast to the developing face processing system, the mature system is less amenable to changes following training and experience (Pascalis et al., 2020). This view is supported by cross-sectional studies in groups of individuals who have unique experiences with faces. For example, White and colleagues found that despite enduring, extensive experience with unfamiliar faces, border control police officers showed poor ability to match such faces, and their performance was no different than that of naïve controls (White, Kemp, Jenkins, Matheson, et al., 2014). Similarly, obstetrician-gynecologist nurses performed poorly and comparably to controls on a task that required identification of newborn faces, despite their extensive experience with these faces (Yovel et al., 2012). Finally, training programs have been shown to be successful in reducing the

ORE in adults (Tanaka & Pierce, 2009), but this effect is specific to perceptual processing and does not appear to extend to memory (McGugin et al., 2011).

In recent years, different groups have demonstrated some level of plasticity even in the matured face processing system. Training programs led to slight improvements in face perception abilities, particularly when observers were asked to individuate faces and received feedback on their decisions (White, Kemp, Jenkins, & Burton, 2014; Yovel et al., 2012). Additionally, adults with congenital prosopagnosia show improved face perception abilities following a 13-week perceptual training program. The improvement was found to persist for at least three months, and there is some evidence for generalization to new faces (Corrow et al., 2019; for a review see DeGutis et al., 2014). Nevertheless, most of the studies described so far are limited in terms of the number of participants, the non-ecological nature of the training regime, and, perhaps most importantly, the length and quality of exposure.

Here, we aimed to overcome these issues by evaluating face perception abilities of masked faces in the era of the COVID-19 pandemic. To this end, we used the well-established Cambridge Face Memory Task (CFMT) (Duchaine & Nakayama, 2006; Russell et al., 2009). Since it was introduced, the CFMT has been used extensively to evaluate face processing capabilities (e.g., Bobak et al., 2016; DeGutis et al., 2013; Wilmer et al., 2010), to identify individuals with superior face processing abilities (CFMT+: Russell et al., 2009), and to identify individuals who suffer from prosopagnosia (Avidan et al., 2011; Duchaine & Nakayama, 2006). Other studies have demonstrated that the CFMT has high reliability (Bowles et al., 2009) and is well-correlated with other face perception tests (DeGutis et al., 2013; Stacchi et al., 2020), with self-rated face recognition ability (Bowles et al., 2009), and, importantly, with naturalistic assessments of face perception abilities (Balas & Saville, 2017). Thus, the CFMT is a well-suited tool for estimating face processing abilities in the era of the pandemic.

We first characterized face recognition abilities for masked and non-masked faces at a critical time-point, just when masks became highly prevalent around the globe (May 2020) (Freud et al., 2020), and continued to track performance at four additional time-points over a period of 16 months. At each time point, we recruited a large sample of participants (N across time-points = 1,473). Importantly, each participant completed the CFMT with upright and inverted faces to measure qualitative changes in the processing of masked faces. Finally, we also conducted a set of control experiments and analyses to estimate whether the amount of individual experience with masked faces can account for the persistent deficit in the recognition of masked faces.

Methods

Participants

A total of 1,473 participants were tested at five different time-points: May 2020 (accuracy data were previously reported (Freud et al., 2020)), September 2020, January 2021, May 2021, and September 2021. All participants were recruited online (https://www.prolific.co/) and were compensated for their time (~$6 CAD for 30 min). All participants were included in the final analysis for the first four time-points. Due to a technical error, participants in September 2021 were allowed to complete several sessions of the experiment. We identified all participants who completed the experiment more than once and discarded their data from the repeated sessions. The large sample size at each time-point was utilized to ensure sufficient statistical power to identify any changes in face perception abilities across time-points. Table 1 includes a summary of the participants' information at each time point. Figure 2 shows the age distribution across time points as well as the number of participants per country of residence. Notably, demographic variables were largely similar across time points. Note that for the last time point (September 2021), there was a large group of participants from South Africa (relative to other time points). To account for this discrepancy, we have repeated the analyses reported below, excluding these participants. Similar results were obtained.

We recruited two additional groups of participants (n=300, September 2021; n=283, October 20210), who completed control experiments that were designed to explore the relationship between reported experience with masks and recognition abilities of masked faces.

Table 1: Demographic details across time-points

	May 2020		Sept 2020		January 2021		May 2021		Sept 2021	
	Masked	Non-Masked	Masked	Non-Masked	Masked	Non-Masked	Masked	Non-Masked	Masked	Non-Masked
N (female)	146 (76)	147 (75)	147 (77)	0	248 (128)	247 (124)	152 (78)	149 (76)	127 (65)	109 (71)
Age (SD)	25.7 (7.7)	25.4 (7.2)	26.4 (8.8)	NA	25.4 (7.6)	27.1 (9.7)	24.8 (6.8)	26.0 (8.4)	27.4 (10.0)	26.7 (7.4)

	Sept 2021 (Within-subject CFMT)	Oct 2021 (GFMT-masked version)	
		Masked	Non-Masked
N (female)	300 (150)	283 (142)	0
Age (SD)	26.3 (7.8)	26.3 (7.8)	NA

All experiments were performed according to relevant guidelines and regulations of the ethics review board at (BLIND COPY). All participants provided informed consent. Data and analysis code for all experiments are available on the Open-Source Framework (BLIND COPY) under CC-By Attribution 4.0 International license.

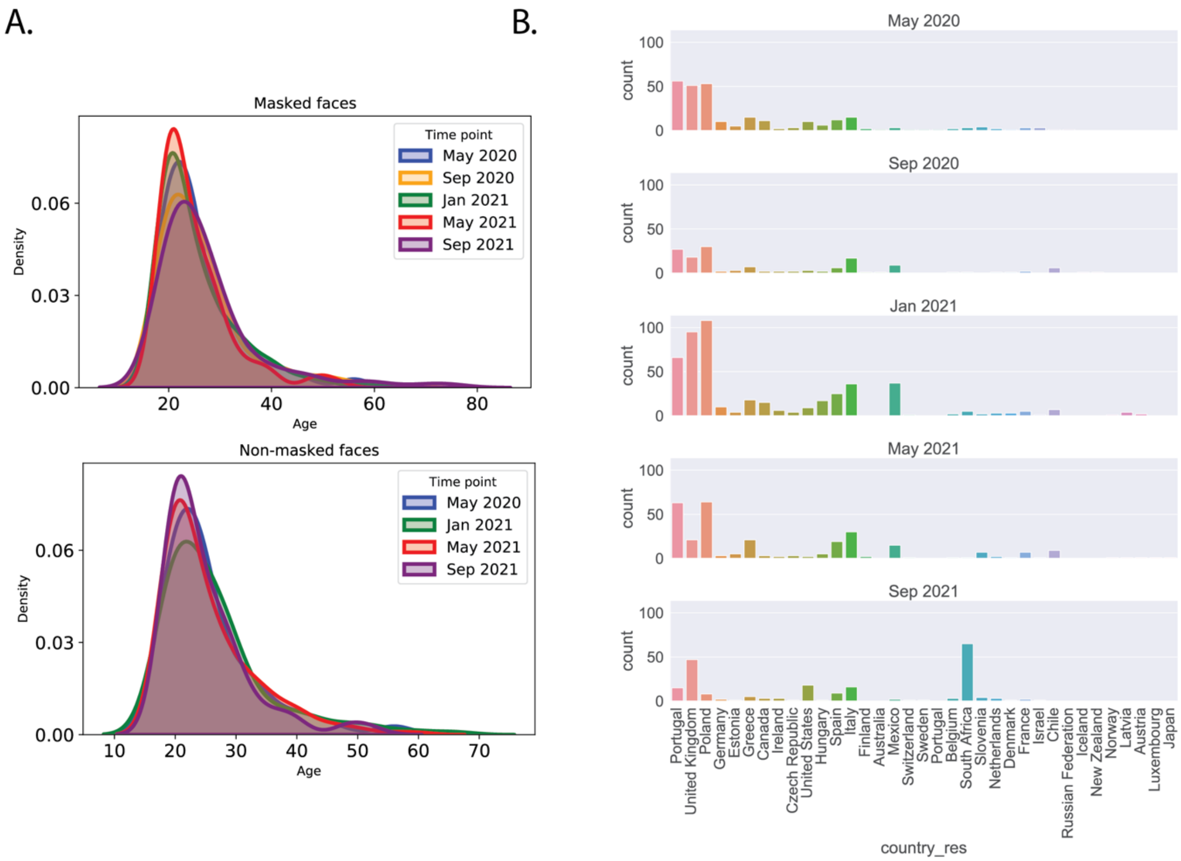

Figure 2: Demographic properties of participants. The demographic properties were similar across the four testing time points. (A) Density histogram of participants' age. (B) Distribution of participants' country of residence.

Materials

The extended version of the CFMT (Duchaine & Nakayama, 2006; Russell et al., 2009) was used to assess face perception abilities. The standard CFMT includes three phases (total of 72 trials) with increasing levels of difficulty. The first phase (easy) involves learning to recognize six unfamiliar male faces from three different viewpoints and then testing recognition of these faces in a three-alternative forced-choice task. The second phase involves a refresher in which the six faces are presented simultaneously from one (frontal) viewpoint followed by testing from novel viewpoints and lighting conditions. The third (difficult) phase is similar to the second phase but includes test images with added visual noise. The long form of the CFMT includes an additional 30 trials with an even higher level of difficulty in which novel images varying in pose, emotional expression, and amount of information available are presented. This latter part is typically used to identify super-recognizers (individuals with extraordinarily high face recognition memory) (Russell et al., 2009), while the first three phases are more sensitive to detecting basic performance and potential deficits within face perception abilities (Murray & Bate, 2020). For the current study, we limited the analyses to the standard form of the CFMT (levels 1-3).

For the last time-point (September 2021) and for the control experiments, we included a questionnaire that measured the Experience with Masked Faces (EMF) (Table 2). The questionnaire includes 4 scales: 1) level of experience that each participant had with masked faces ("Experience" scale – items 1, 2,7,8), 2) extent to which regulations about mask wearing were enforced in the participant's country of residence ("Regulation" scale – items 3,4), 3) subjective difficulty in recognizing masked faces ("Subjective" scale – items 5,6), And 4) continuous evaluation of experience with masked faces (item 9).

Table 2: Experience with Masked Faces Questionnaire (List of items)

Please rate the degree to which you agree or disagree with the following eight statements: 1 (strongly disagree) to 7 (strongly agree)

1. *I have a lot of experience in encountering people who wear masks*
2. *I spend a lot of time in public places where mask wearing is common*
3. *Mask wearing is strongly enforced in my country of residence*
4. *Most people follow mask mandates in my country*

5. *I have difficulties recognizing familiar people (i.e., family members, colleagues, friends) when they are wearing masks*
6. *When people wear masks, I have difficulties in deciding whether they are familiar to me or not*
7. *Most people who I encounter everyday are wearing masks*
8. *In my daily life, I need to recognize a lot of people who wear masks*
9. *What is the proportion of people you know who you encounter wearing a mask (0%-100%)?*

We also asked participants to indicate whether they work from home, remotely, or are unemployed. However, upon analyzing the data, we noted that this question was confounded with the participant's age (remote workers were generally older than the other groups). Thus, we did not use this measure in any additional analyses.

Lastly, in October 2021, also collected data using a different paradigm in which external face cues were available. In particular, we used the short version of the Glasgow Face Matching Test (GFMT) (Burton et al., 2010). In this task, participants are shown 40 pairs of faces, photographed in full-face view but with different cameras, and are asked to make same/ different judgments. Masks were graphically added to all faces.

Procedure

For the main experiment, participants were randomly assigned to one of two groups. The first group completed the original CFMT (faces without masks), while the second group completed a modified version of the CFMT in which an identical face mask was added to all faces (Figure 1). To explore the processing style of faces with and without masks, each participant completed the test twice, once with upright faces and once with inverted faces. Block order (upright or inverted) was counterbalanced between participants. Due to a technical error, for the September 2020 testing point, only one order was employed (upright faces followed by inverted faces). Accuracy scores (0-72) for upright and inverted faces were computed and served as the main dependent variable. We also analyzed reaction times (RTs) for correct trials. For the control experiments, participants completed the masked and non-masked versions of the CFMT (only the upright version) or the masked version of the GFMT. Statistical analyses were conducted using JASP (JASP team, 2020) and in-house codes written in Python.

Results

We explored the extent to which persistent exposure to masked faces facilitated recognition abilities of these faces. To this end, participants completed the CFMT (Duchaine & Nakayama, 2006) with upright and inverted faces (within-subjects) while the faces were either masked or non-masked (between-subjects) across five different time-points (between-subjects) over a period of 16 months, during which face mask wearing became prevalent on a daily basis because of the COVID-19 pandemic.

Figure 3A shows the group averages across conditions on the standard CFMT for each of the five time points. The results show a persistent deficit in face recognition abilities for masked faces. In addition, for all time points, the face inversion effect was reduced for masked faces, pointing to a qualitative difference in the processing of these faces, which was again persistent over time.

To statistically validate these results, we first employed ANOVA with time points, orientation, gender, and face type (masked/non-masked). Note that for this analysis, we included only four time points, as we do not have data for the non-masked faces condition for September 2020. The ANOVA demonstrated main effects of group [non-masked better than masked; $F_{(1,1310)}=174$, $p<.001$, $\eta^2_p = 0.12$], orientation [upright > inverted; $F_{(1,1310)}=2307$, $p<.001$, $\eta^2_p = 0.64$], and gender [females > males; $F_{(1,1310)}=48.2$, $p<.001$, $\eta^2_p = 0.03$]. An interaction was found between orientation and group, with a reduced inversion effect for masked faces [$F_{(1,1310)}=202$,

$p<.001$, $\eta^2_p = 0.13$] that was consistent across time points [three-way interaction: $F < 1$]. The reduced inversion effect could not be easily accounted for by a floor effect for the inverted masked faces, as the average CFMT score was well above chance level (33%) even for these faces. Importantly, we did not find an effect of time point [$F_{(3,1310)}=2.513$, $p=.057$, $\eta^2_p = 0.006$] nor an interaction between time point and group [$F_{(3,1310)}=1.561$, $p=.197$, $\eta^2_p = 0.004$]. Similar results were observed when participants' age served as a covariate. These results provide evidence for a persistent decrement in face perception abilities across time points, accompanied by a qualitative change in the processing of the masked faces.

Next, to further evaluate the consistency of recognition performance of masked faces across time points, we focused our analysis specifically on the upright masked faces condition. We conducted a Bayesian ANOVA with time point (five levels) on the accuracy scores. In contrast to Null Hypothesis Significance Testing, a Bayesian ANOVA can also provide evidence in favor of the null hypothesis (Wagenmakers et al., 2018). Accordingly, the Bayesian ANOVA decisively supported the null hypothesis (i.e., no difference between the five time points) [$BF_{10} = 0.022$], such that the null hypothesis was 166 times more likely than the H1 hypothesis. Similar results were observed when the Bayesian ANOVA was employed on the masked inverted faces [$BF_{10} = .008$]. It worth mentioning that even numerically, the CFMT score for masked faces was lowest for the last collected datapoints (May and September 2021).

Finally, given the naturalistic nature of our sampling methods, some demographic differences were observed across the different samples. To validate that our results could not be accounted by these differences, we conducted an additional, matched analysis of the data. In particular, we matched participants across the different time points based on their country of residence, gender, and age (±5 years). The new matched sample consists of 41 participants per time point (only for the masked condition). We were able to replicate the findings reported above – that is, consistent performance for masked faces with no improvement over time [Upright: BF_{10}=0.062; Inverted: BF_{10}=0.064].

It is notable that the large sample size allowed us to compare not only the means but also the distribution of results across individuals at each of the time points. We focused this analysis on the upright orientation separately for masked and non-masked faces. As presented in Figure 3B, we did not observe any significant changes in the distribution of the results across time points for both the masked and the non-masked conditions. Two-sample Kolmogorov–Smirnov tests were employed across possible combinations for the masked and non-masked conditions, and confirmed that the distributions were not different from each other at the different time points [$Ds < 0.2$, $p > .2$]. Together, the analyses of the accuracy scores provide robust evidence against changes in the processing style or processing efficiency of masked faces following extended exposure.

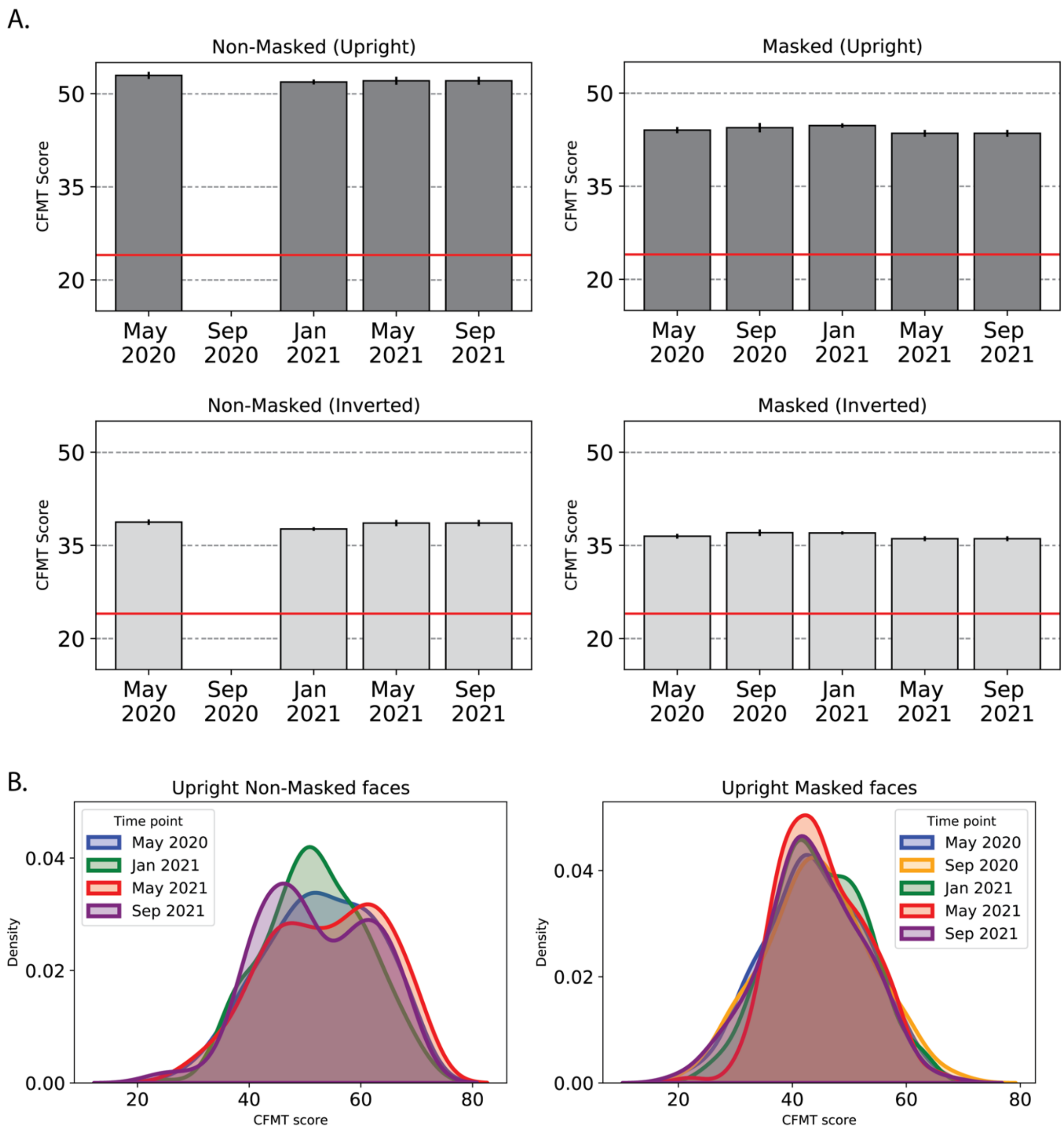

Figure 3: Accuracy results across time points. (A) Results of the standard CFMT experiment for non-masked and masked faces. For all time points, performance was significantly impaired for masked compared to non-masked faces. An inversion effect was found for both face types, but it was significantly smaller for masked faces. The results are similar across the different time points. Note that in September 2020, data was collected only for the masked faces. Error bars represent the standard error of the mean. The red horizontal line shows the chance level score. (B) Distribution of the CFMT results across the different time points, separately for upright masked and non-masked faces. Similar distributions were observed across the different time points. The results show that extended exposure to masked faces did not alter their recognition.

Reaction times (RT) are not commonly analyzed for the CFMT. However, we chose to analyze the RTs to provide an additional measure for face perception abilities and to account for possible speed-accuracy trade-off effects. As shown in Figure 4, the RT results mainly mirrored the effects observed for the accuracy scores. In particular, the ANOVA demonstrated main effects of orientation [upright faster than inverted; $F_{(1,1310)}=77.8$, $p<.001$, $\eta^2_p = 0.056$] and gender [females faster than males; $F_{(1,1310)}=7.552$, $p<.05$, $\eta^2_p = 0.006$]. The interaction between orientation and group was also significant, with a reduced inversion effect for masked faces [$F_{(1,1310)}=30.27$, $p<.001$, $\eta^2_p = 0.02$] despite the lack of a main effect for group [$F_{(1,1310)} < 1$]. There was a weak main effect for time-point, with longer reaction times for the last time point (September 2021) [$F_{(3,1310)}=5.01$, $p<.01$, $\eta^2_p = 0.01$]. Importantly, there was no interaction between time point and group (mask status) [$F_{(1,1310)} < 1$] on reaction times, suggesting that the mask effect remained constant across time points. Finally, a Bayesian ANOVA on the RT for upright faces did not provide evidence for neither the null hypothesis nor the H1 hypothesis [$BF_{10} = .89$]. Notably, however, the RTs for the last sample were in fact longer than any other time point, refuting the notion that participants became more efficient in the processing of masked faces. Thus, the RT results support the conclusion that time and experience with masked faces did not change or improve the face mask effect.

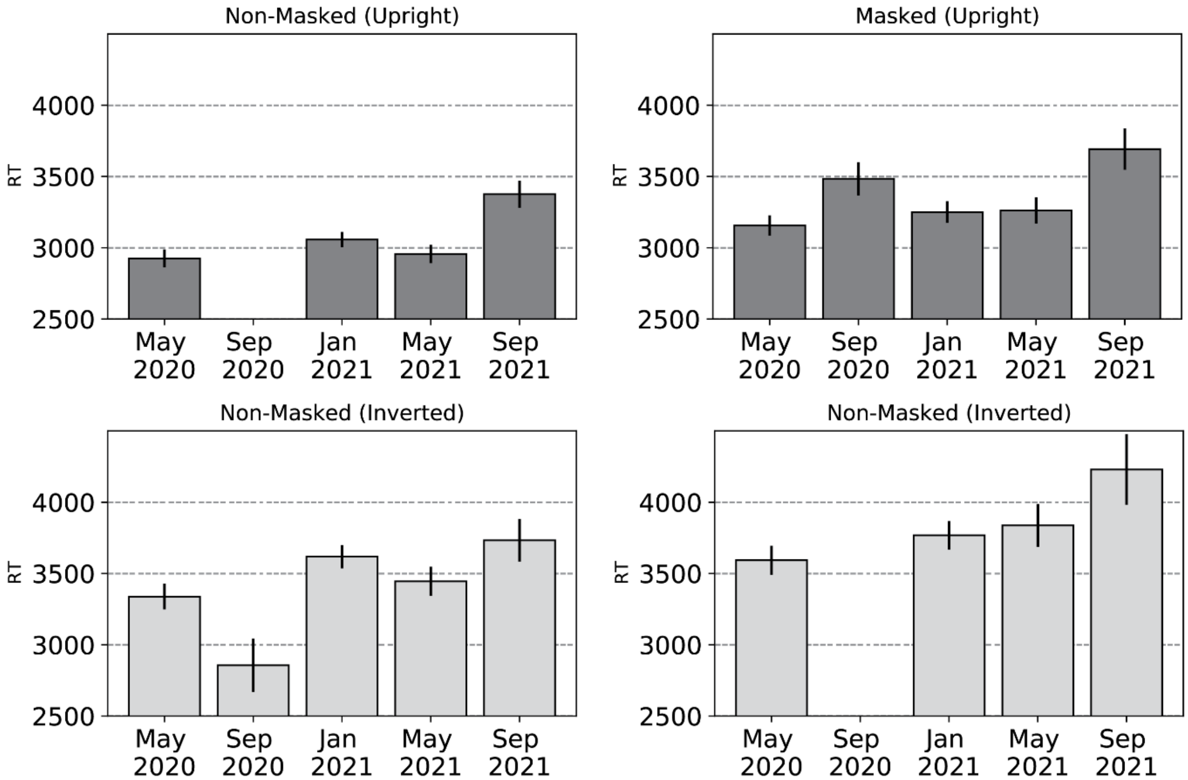

Figure 4: RT results across time points. (A) Results of the standard CFMT experiment for non-masked and masked faces. Across all time points, performance was significantly impaired for masked compared to non-masked faces. An inversion effect was found for both face types, but it was significantly smaller for masked faces. The results are similar across the different time-points. Note that in September 2020, data were collected only for the masked faces. Error bars represent the standard error of the mean.

Individual differences with masked faces

So far, we demonstrated that the mask effect was consistent across time-points. However, it is difficult to draw strong conclusions about the absence of perceptual learning without knowing the characteristics of the perceptual experiences of the target populations. Specifically, as the pandemic continued, some people (e.g., health workers) had progressively gained greater experience with masked faces, while others (e.g., students who learned from home) did not gain the same amount of experience. Thus, it is plausible that we were unable to identify significant changes in the mask effect because only a sub-group of our participants showed this type of improvement. Although this conclusion is not supported by the averaged data, which should still show a directional shift toward better performance with masked faces over time, we decided to further address this concern using an additional set of experiments and analyses aimed at testing the possible role of individual differences in exposure to masked faces.

The "Experience with Masked Faces" (EMF) Questionnaire

First, we reached out to participants who completed the experiment in September 2021 and asked them to fill out the EMF questionnaire to evaluate their experience with masked faces (see Methods). 200 (out of 236) participants completed the questionnaire. The following analysis was conducted only on the participants who completed the "mask" condition (N=110), but similar results were obtained when both groups were included.

We found a robust correlation between the Experience scale (items 1,2,7,8) and the continuous evaluation of experience with masked faces (item 9) [$r=.64$, $p < .001$], between the Experience and Regulation scales [$r=.54$, $p < .001$], and between the Regulation scale and continuous evaluation of experience with masked faces [$r=.62$, $p < .001$]. As expected, we also found strong correlations between performance with upright and inverted faces for both accuracy and RT. Importantly, we did not find any correlations between face perception abilities and experience with masked faces. This was true for accuracy scores and RT for both upright and inverted faces [all $ps > .1$] (Figure 5A).

These results indicate that the amount of reported experience with masked faces is not correlated with performance. Thus, the lack of improvement in recognition of masked faces is unlikely to reflect a reduced amount of experience with these faces.

A.

	Upright	Inverted	RT upright	RT inverted	Experience	Regulations	Subjective	% of Masked Faces
Upright								
Inverted	0.47*							
RT upright	-0.07	0.01						
RT inverted	-0.16	-0.11	0.36*					
Experience	0.02	0.07	0.07	0.09				
Regulations	0.04	0.06	0.09	0.14	0.57*			
Subjective	0.02	-0.03	-0.05	0.04	0.00	0.06		
% of Masked Faces	0.10	-0.01	0.06	0.13	0.66*	0.66*	-0.01	

B.

	Mask Effect	Mask Effect RT	Experience	Regulations	Subjective	% of Masked Faces
Mask Effect						
Mask Effect RT	-0.00					
Experience	-0.03	0.02				
Regulations	0.10	0.03	0.41*			
Subjective	-0.07	-0.09	0.02	0.09		
% of Masked Faces	-0.08	-0.04	0.64*	0.46*	0.03	

C.

	d'	RT	Experience	Regulations	Subjective	% of Masked Faces
d'						
RT	0.22*					
Experience	-0.06	0.06				
Regulations	-0.01	0.20*	0.51*			
Subjective	-0.06	-0.01	-0.01	0.04		
% of Masked Faces	-0.04	0.16*	0.65*	0.54*	-0.02	

Figure 5: Correlations between face perception measurements and the EMF scales. (A) September 2021 time-point – upright and inverted masked faces. (B) CFMT – masked and non-masked were included as a within-subject variable, and a mask effect was calculated for accuracy and RTs. (C) GFMT (short version) for masked faces. d' (sensitivity) and RT were calculated. Across the different experiments, high correlations were found between the Experience scale, the Regulation scale and the % of masked faces question, reflecting the reliability of these items. No correlations were found with face perception measurements,

suggesting that the amount of experience with masked faces is not correlated with individual differences in masked face recognition. Significant correlations ($p < .05$) are denoted by asterisks.

CFMT – within-subject design

The results of the previous analysis showed that there were no correlations between the CFMT scores and the amount of reported experience with masked faces. This was true for participants who completed the masked version of the CFMT and for those who completed the non-masked version of the CFMT. Nevertheless, because we used a between-subject design, we could not directly examine whether the mask effect was modulated by the amount of experience with masked faces. Thus, we recruited a new group of participants who completed the CFMT (upright version) twice – with masked and non-masked faces - as well as the EMF questionnaire.

A repeated-measures ANOVA with mask status as a within-subject variable revealed a robust main effect [$F_{(1,299)}=499$, $p<.001$, $\eta^2_p = 0.626$], with greater accuracy for non-masked faces [CFMT non-masked: 54.4; CFMT masked: 44.81; Mask effect = ~18%]. RT analysis corroborated this finding with longer RTs for masked faces [$F_{(1,299)}=19.2$, $p<.001$, $\eta^2_p = 0.06$; CFMT non-masked: 3631ms ; CFMT masked: 4539ms].

Next, we examined the relationship between the mask effect (in terms of accuracy and RT) and the scores obtained from the EMF questionnaire (Figure 5B). As expected, we found a robust correlation between the Experience scale (items 1,2,7,8) 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 3,4) scales [$r=.41$, $p < .001$] and between the Regulation scale and continuous evaluation of experience with masked faces [$r=.46$, $p < .001$]. Importantly, we did not find any correlations between the mask effect (either in terms of accuracy or RT) and experience with masked faces [all ps >.1].

Finally, 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 of masked face experiences (students and unemployed, n=124). First, as excepted, we found that health and retail workers had more experience with masked faces [Experience scale: $F_{(3,166)}=8.33$, $p<.001$, $\eta^2_p = 0.13$; % of Masked Faces scale: $F_{(3,166)}=8.54$, $p<.001$, $\eta^2_p = 0.13$]. Critically, an additional ANOVA with workplace and mask status showed a robust main effect for mask status [$F_{(1,166)}=22.3$, $p<.001$, $\eta^2_p = 0.119$], but no main effect for workplace [$F < 1$] nor an interaction between these factors [$F < 1$], suggesting that the mask effect was not modulated by the participants' occupational profile.

Together, the results of the within-subject experiment reinforce the notion that the lack of improvement across time-points does not reflect individual differences in the amount of exposure to masked faces, but rather the relatively rigid nature of matured face processing mechanisms.

Glasgow Face Match Test (GFMT) - masked faces

Last, we collected data from an additional established test – the GFMT (Burton et al., 2010). In the short GFMT, participants are shown 40 pairs of faces, photographed in full-face view but with different cameras, and are asked to make same/ different judgments. One advantage of the GFMT over the CFMT is that external face cues (e.g., hair) that might assist in identification of masked faces are not removed. Previous research has already demonstrated the existence of the mask effect for this task (Carragher & Hancock, 2020), but it is still unclear whether experience with masked faces is correlated with performance.

To test the possible role of experience, we examined the relationship between face perception performance (d' and RT) and the scores obtained from the EMF questionnaire (Figure

5C). Similar to the results observed for the CFMT (see above), we found a robust correlation between the Experience scale (items 1,2,7,8) 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 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 continuous evaluation of experience with masked faces [$r=.16$], such that particpants who reported greater experince with masked faces had slower reaction times. Note that a reversed pattern is predicted if greater exposure to masked faces improves processing efficency.

Discussion

Face masks were an important tool in the effort to minimize COVID-19 virus transmission (Cheng et al., 2020). Accordingly, the years of 2020-2021 provided an unprecedented opportunity to examine the effects of prolonged and frequent exposure to occluded faces on recognition abilities. Here, we have documented persistent quantitative and qualitative alterations in face processing abilities for masked vs. non-masked faces, with no evidence of changes in the processing of masked faces over time. We found that the CFMT scores for upright faces decreased by ~15% when masks were added to the faces. This reduction remained statistically constant across a 16-month period, as verified by a Bayesian analysis. This finding suggests that the matured face processing system did not benefit from the prolonged exposure to masked faces. Additional experiments and analyses showed that the consistent decrement in face processing of masked faces was evident even when individual difference in exposure to these faces were considered.

Another key finding is the consistent and robust reduction of the face inversion effect for masked faces across all time points. In particular, the inversion effect was roughly 43% smaller for masked faces. This modulation of the inversion effect was evident not only in terms of accuracy, but also in terms of reaction time. The inversion effect is suggested to reflect a failure to extract the configural relationships between face parts (Farah et al., 1995; Freire et al., 2000). Hence, the smaller inversion effect for masked faces may be taken as evidence that holistic processing is largely reduced (though not entirely abolished). This qualitative change in the processing of masked faces was consistent across time points, providing additional evidence for the relative rigidity of the matured face processing system.

Rigidity of the matured face perception system

Face perception rapidly develops in infancy but is then subject to a prolonged developmental trajectory (Pascalis et al., 2011, 2020). In early childhood, face processing is shaped by experience with other faces (Bate et al., 2020). One of the best examples of this malleability comes from the ORE, which is evident early in life (Kelly et al., 2009) but could be reversed or disappear if a child is regularly exposed to other-race faces (De Heering et al., 2010; Sangrigoli et al., 2005). This plasticity might reflect a more general property of the developing visual system and has been found even in clinical populations with extensive brain injuries (Liu et al., 2019; Mundinano et al., 2019).

In contrast, in adulthood, face processing mechanisms are already in place and are less likely to be affected by experience (Pascalis et al., 2020; White, Kemp, Jenkins, Matheson, et al., 2014; Yovel et al., 2012). Here, we show that even extensive, daily, naturalistic encounters with masked faces do not facilitate the recognition of these faces. This finding provides support for the idea that the face processing system becomes relatively rigid in adulthood and is therefore unlikely to be tuned by specific experiences.

Notably, however, several attempts to improve face perception in adulthood have been moderately successful, even among clinical population (DeGutis et al., 2014). The common aspect of these attempts was the inclusion of an individuation task (McGugin et al., 2011; Yovel et al., 2012) and ongoing feedback (White, Kemp, Jenkins, & Burton, 2014). Those two elements are not consistently provided in our everyday experience with masked faces, as it is often the case that masked faces (in the grocery store, on public transportation, etc.) are not identified, and observers do not receive feedback. Given the importance of these elements to improvement in face perception abilities in adulthood, future studies should examine whether a training program that includes consistent feedback and an individuation process could reduce the observed mask effect. It would also be interesting to examine whether such improvements will be accompanied by an increased inversion effect.

Limitations

The current investigation is timely and unique and benefits from the large sample size. However, there are still important limitations that should be acknowledged and addressed in future studies. First, while the CFMT is a reliable test that has been used extensively over the last two decades (Bobak et al., 2016; Russell et al., 2009), the faces included in this test are all Caucasian males. Given the gender effect observed in our data, as well as by other groups (Bobak et al., 2016), it is important to examine the reported effects using other, more diverse tests (Scherf et al., 2017). Another concern regards the ecological validity of the CFMT. Specifically, external face cues, which are important for real-life face recognition, are not available in this test. This concern might even be more detrimental in the case of masked faces. However, it is important to note that previous studies reported on correlations between CFMT scores and subjective reports of face recognition abilities (Shah et al., 2015), between the CFMT and other measurements of face processing abilities (DeGutis et al., 2013; Russell et al., 2009), and, most importantly, between CFMT scores and naturalistic assessments of face perception abilities (Balas & Saville, 2017). It is also worth noting that previous studies demonstrated the existence of the mask effect for other test and image sets such that GFMT (Carragher & Hancock, 2020 and also see the control experiment described above) and the KDEF (Marini et al., 2021) in which external face cues are preserved. The concern regarding the ecological validity could also be extended to the fact that across the experiments, faces did not include other cues that might facilitate person recognition such as motion, voice and body-shape. Importantly, however, it is established that faces have a superior role in person recognition even when other cues are available (Hahn et al., 2016). This conclusion is also demonstrated in cases prosopagnosia, which is experienced in daily life even when all cues are available.

Another limitation of the current image set (as well as other image sets utilized by previous studies) is that the masks were added in an artificial manner to existing pictures. This might lead to an omission of face shape cues that are normally available and plausibly critical in recognizing masked faces in naturalistic settings. While we cannot rule out the detrimental effect of the artificial mask on face perception, a recent study by Marini and colleagues (2021) demonstrated the existence of a mask effect even for transparent masks that reveal important cues from the lower part of the face. Hence, it is unlikely to assume that the mask effect observed here, and, in particular, the lack of improvement in face perception for masked faces, could be attributed solely to the nature of the stimuli.

Finally, we adopted a cross-sectional approach that, in contrast to a longitudinal approach, does not allow us to examine within-subject changes. Nevertheless, the cross-sectional approach has several important advantages. First, we were able to recruit a large group of participants with no attrition. Second, and, most important, each participant completed the experiment only once, and therefore there was no risk of priming effects that reflect experience with the CFMT and/or the stimulus set.

Conclusion

The current study provides evidence for a persistent deficit in recognizing masked faces, despite extensive, naturalistic exposure to these faces. Deficient performance in recognizing masked faces, along with the qualitative difference in processing those faces, may have long-term implications for daily activities, especially social interactions. In addition, continuous exposure to masked faces could also lead to long-lasting effects in other domains, such as processing of facial emotion and to a prolonged feeling of anxiety and alienation in a world heavily occupied by masked people (Calbi et al., 2021). The current study also provides important insights into the limited malleability of the matured face processing system and can, therefore, inform future efforts to enhance or recover visual abilities in adulthood.

References

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404–1412. https://doi.org/10.1038/nn.4635

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia*. https://doi.org/S0028-3932(11)00236-3 [pii] 10.1016/j.neuropsychologia.2011.05.002

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research*, *141*, 228–236.

Bate, S., Adams, A., & Bennetts, R. J. (2020). Guess who? Facial identity discrimination training improves face memory in typically developing children. *Journal of Experimental Psychology: General*, *149*(5), 901.

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, *7*, 1378.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C. E., & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Calbi, M., Langiulli, N., Ferroni, F., Montalti, M., Kolesnikov, A., Gallese, V., & Umiltà, M. A. (2021). The consequences of COVID-19 on social interactions: An online study on face covering. *Scientific Reports*, *11*(1), 1–10.

Carragher, D. J., & Hancock, P. J. B. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cognitive Research: Principles and Implications*, *5*(1), 59. https://doi.org/10.1186/s41235-020-00258-x

Cheng, V. C.-C., Wong, S.-C., Chuang, V. W.-M., So, S. Y.-C., Chen, J. H.-K., Sridhar, S., To, K. K.-W., Chan, J. F.-W., Hung, I. F.-N., & Ho, P.-L. (2020). The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *Journal of Infection*, *81*(1), 107–114.

Chung, M.-S., & Thomson, D. M. (1995). Development of face recognition. *British Journal of Psychology*, *86*(1), 55–87.

Corrow, S. L., Davies-Thompson, J., Fletcher, K., Hills, C., Corrow, J. C., & Barton, J. J. (2019). Training face perception in developmental prosopagnosia through perceptual learning. *Neuropsychologia*, *134*, 107196.

De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, *13*(1), 181–187.

DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, *8*, 561.

DeGutis, J. M., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Dhamecha, T. I., Singh, R., Vatsa, M., & Kumar, A. (2014). Recognizing disguised faces: Human and machine evaluation. *PloS One*, *9*(7), e99212.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What Causes the Face Inversion Effect. *Journal of Experimental Psychology-Human Perception and Performance*, *21*, 628–634.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*(2), 159–170.

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports*, *10*(1), 1–8.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Grundmann, F., Epstude, K., & Scheibe, S. (2021). Face masks reduce emotion-recognition accuracy and perceived closeness. *PLOS ONE*, *16*(4), e0249792. https://doi.org/10.1371/journal.pone.0249792

Hahn, C. A., O'Toole, A. J., & Phillips, P. J. (2016). Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*, *107*(1), 117–134.

JASP team. (2020). *JASP* (0.14) [Computer software].

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Blackwell.

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105–114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31–F36. https://doi.org/10.1111/j.1467-7687.2005.0434a.x

Kret, M., & De Gelder, B. (2012). Islamic Headdress Influences How Emotion is Recognized from the Eyes. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00110

Liu, T. T., Freud, E., Patterson, C., & Behrmann, M. (2019). Perceptual function and category-selective neural organization in children with resections of visual cortex. *Journal of Neuroscience*, *39*(32), 6299–6314.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Marini, M., Ansani, A., Paglieri, F., Caruana, F., & Viola, M. (2021). The impact of facemasks on emotion recognition, trust attribution and re-identification. *Scientific Reports*, *11*(1), 5577. https://doi.org/10.1038/s41598-021-84806-5

Maurer, D. (2017). Critical periods re-examined: Evidence from children treated for dense cataracts. *Cognitive Development*, *42*, 27–36.

McGugin, R. W., Tanaka, J. W., Lebrecht, S., Tarr, M. J., & Gauthier, I. (2011). Race-Specific Perceptual Discrimination Improvement Following Short Individuation Training With Faces. *Cognitive Science*, *35*(2), 330–347. https://doi.org/10.1111/j.1551-6709.2010.01148.x

McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., & Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports*, *9*(1), 1–13.

Mundinano, I.-C., Chen, J., de Souza, M., Sarossy, M. G., Joanisse, M. F., Goodale, M. A., & Bourne, J. A. (2019). More than blindsight: Case report of a child with extraordinary visual capacity following perinatal bilateral occipital lobe injury. *Neuropsychologia*, *128*, 178–186.

Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: Repeat assessment using the Cambridge Face Memory Test. *Royal Society Open Science*, *7*(9), 200884. https://doi.org/10.1098/rsos.200884

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*(1), 79–85.

Pascalis, O., Fort, M., & Quinn, P. C. (2020). Development of face processing: Are there critical or sensitive periods? *Current Opinion in Behavioral Sciences*, *36*, 7–12. https://doi.org/10.1016/j.cobeha.2020.05.005

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, *102*(14), 5297–5300.

Pascalis, O., Viviés, X. de M. de, Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing. *WIREs Cognitive Science*, *2*(6), 666–675. https://doi.org/10.1002/wcs.146

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257.

Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440–444.

Scherf, K. S., Elbich, D. B., & Motta-Mena, N. V. (2017). Investigating the influence of biological sex on the behavioral and neural basis of face recognition. *Eneuro*, *4*(3).

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–729.

Scott, L. S., & Monesson, A. (2009). The Origin of Biases in Face Perception. *Psychological Science*, *20*(6), 676–680. https://doi.org/10.1111/j.1467-9280.2009.02348.x

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): A self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 140343.

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive Research: Principles and Implications*, *5*(1), 1–17.

Stajduhar, A., Ganel, T., Avidan, G., Rosenbaum, R. S., & Freud, E. (2021). *Face Masks Disrupt Holistic Processing and Face Perception in School-Age Children*.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 122–131.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., & Boutin, B. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76.

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100–106.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport Officers' Errors in Face Matching. *PLOS ONE*, *9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 285–289. https://doi.org/10.1037/a0027077