Journal of Experimental Child Psychology 232 (2023) 105676

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Robust holistic face processing in early childhood during the COVID-19 pandemic

Tristan S. Yates [a,*], David J. Lewkowicz [b,c]

[a] Department of Psychology, Yale University, New Haven, CT 06520, USA
[b] Haskins Laboratories, New Haven, CT 06511, USA
[c] Yale Child Study Center, Yale University School of Medicine, New Haven, CT 06520, USA

A R T I C L E   I N F O

Article history:
Received 6 January 2023
Revised 2 March 2023
Available online 10 March 2023

Keywords:
Early childhood
Composite face effect
Face recognition
Experience
Perception
Online Testing

A B S T R A C T

The timing of the developmental emergence of holistic face processing and its sensitivity to experience in early childhood are somewhat controversial topics. To investigate holistic face perception in early childhood, we used an online testing platform and administered a two-alternative forced-choice task to 4-, 5-, and 6-year-old children. The children saw pairs of composite faces and needed to decide whether the faces were the same or different. To determine whether experience with masked faces may have negatively affected holistic processing, we also administered a parental questionnaire to assess the children's exposure to masked faces during the COVID-19 pandemic. We found that all three age groups performed holistic face processing when the faces were upright (Experiment 1) but not when the faces were inverted (Experiment 2), that response accuracy increased with age, and that response accuracy was not related to degree of exposure to masked faces. These results indicate that holistic face processing is relatively robust in early childhood and that short-term exposure to partially visible faces does not negatively affect young children's holistic face perception.

* Corresponding author.
E-mail address: tristan.yates@yale.edu (T.S. Yates).

https://doi.org/10.1016/j.jecp.2023.105676

Introduction

Mature face processing is based on a perceptual sensitivity to the *holistic* properties of faces. This means that when adults look at faces they respond primarily to the spatial relations among the three most prominent features of a face—namely the eyes, nose, and mouth—and that they glue these features into a gestalt while largely ignoring the specific perceptual attributes associated with each individual feature (Maurer et al., 2002). Mature face processing develops gradually out of some initial perceptual biases that can be observed right at birth (Pascalis et al., 2011; Simion et al., 2007). In essence, infants begin life with two general biases that lead to a preference for face-like stimuli. The first is a bias for the structural properties that characterize face-like and non-face-like objects, and the second is a bias for more elements in the upper part rather than the bottom part of a geometrical stimulus (Simion & Di Giorgio, 2015). Given these two biases, newborns exhibit a general preference for abstract face-like stimuli (Turati et al., 2002), for non-face-like stimuli that exhibit these structural properties (Simion et al., 2002), and for faces themselves (Macchi Cassia et al., 2004).

These initial preferences observed at birth provide a foundation for the gradual emergence of face-specific responsiveness that, to a large extent, is driven by infants' everyday experiences (Pascalis et al., 2020). Infants begin to prefer faces over scrambled face configurations containing more elements in the upper part by 3 months of age (Turati et al., 2005), detect gender differences by 6 months of age (Quinn et al., 2008), and detect facial affect by 8 months of age (Walker-Andrews, 1997). The *experience-dependent* nature of face-specific expertise is illustrated by the fact that newborn infants do not initially prefer or discriminate own- versus other-race (or species) faces, but do just a few months later (Bar-Haim et al., 2006; Kelly et al., 2007; Pascalis et al., 2002). Considering together the empirical evidence on the development of face processing in infancy, it becomes clear that two parallel and concurrent developmental processes lead to the initial growth of face processing expertise in infancy. The first consists of an increasing sensitivity to various aspects of faces, and the second consists of a gradual narrowing from an initially broad sensitivity to potentially socially-relevant information to a more restricted sensitivity to only those categories that are statistically most frequent in infants' everyday environments (Lewkowicz, 2014; Maurer & Werker, 2014).

Importantly, experience-dependent effects on the development of face processing expertise extend well into early childhood. For example, adults who were born in Korea and then adopted by European families in France when they were between 3 and 9 years of age can identify White faces better than Asian faces (Sangrigoli et al., 2005). This demonstrates that the plasticity initially observed in infancy that enables infants to incorporate the statistics of faces into their everyday environment continues into early childhood (Maurer et al., 2005). This, along with findings that face processing expertise continues to grow well into adolescence (Mondloch et al., 2002), makes it theoretically reasonable to hypothesize that everyday experience may have an effect on holistic face processing in early childhood. Despite this possibility, no studies to date have investigated whether everyday experience in visually typical children influences the development of holistic face processing.

Overall, findings have provided a rather mixed picture on the specific age of the developmental emergence of holistic face processing. On the one hand, one study found that holistic face processing emerges as early as 3 months of age (Turati et al., 2010). Moreover, it was reported that a lack of patterned visual input early in life has a lasting impact on holistic face processing into adulthood (Le Grand et al., 2004). On the other hand, other studies found that holistic face processing emerges at 3 to 4 years of age (Crookes & McKone, 2009; de Heering et al., 2007; Macchi Cassia et al., 2009), at 6 years of age (Carey & Diamond, 1994; Mondloch et al., 2007; Ventura et al., 2018), or even as late as 10 years of age (Mondloch et al., 2002). Taken together, this evidence makes it difficult to ascertain precisely when holistic face processing first emerges in development. One possible reason for the divergent estimates of its emergence may be the use of many different paradigms (e.g., face inversion, eye tracking, the part–whole task, the composite face task used here) that measure correlated but distinct aspects of holistic face processing (Boutet et al., 2021). Even within the composite face task, the specific methods of presenting composite faces may influence measures of holistic face processing (Ventura et al., 2018). Given this, we believed that the most conservative approach was to remain agnostic with specific regard to the developmental timing of the emergence of holistic face processing and to simply settle on one of the established experimental methods for testing for it. Once we settled on the composite face technique, we were in position to investigate the effects of altered experience with faces (i.e., experience with partially visible faces) on the development of holistic face processing (Carnevali et al., 2022) while being fully cognizant of the constraints that this specific experimental method places on the ultimate interpretation of our findings.

The aim of the current study was to determine whether everyday experience in early childhood might affect holistic face processing. If the sensitive period for holistic processing extends into early childhood, and if such processing depends on exposure to fully visible faces, it is reasonable to expect that exposure to partially visible faces may have detrimental effects on holistic face perception in young children. Alternatively, exposure to partially visible faces may encourage children to focus more on specific portions of the face, including the eyes, which has been shown to enhance holistic face processing in adults (Wang et al., 2019, 2023). Finally, if holistic face processing is already in place as early as the first year of life (Le Grand et al., 2004; Turati et al., 2010), we may expect to find no influence of differential face input on holistic processing in early childhood. To test these possibilities, we took advantage of the "natural" experiment created by the COVID-19 pandemic when face masks were mandated by public health officials to prevent the spread of the virus. One consequence of these mandates was that, overall, children were exposed more to the top halves of other people's faces (see Table S1 in online supplementary material). That is, even though children continued to see fully visible faces of family members during the initial lockdowns, once lockdowns were lifted and they returned to daycare, preschool, and/or kindergarten, children were exposed to masked faces. Of course, it is difficult to quantify the degree of exposure to masked faces (in terms of both frequency and duration) that might actually be detrimental to holistic face processing. As a result, in the current study we adopted an exploratory approach to investigating the relation between children's exposure to masked/partially visible faces and holistic processing.

Clues as to whether masks might impede children's learning and representation of faces may be gleaned from studies that have assessed the effects of masks and occlusion on face processing. One study found that adults' face processing is disrupted by sunglasses or masks (Noyes et al., 2021), whereas another found that children's holistic face processing is altered (even more so than in adults) when viewing masked individuals (Stajduhar et al., 2022). These findings suggest that children whose typical experience with faces (which usually consists of seeing fully visible faces) is altered due to exposure to partially visible faces for considerable parts of their day may find it difficult to discriminate faces in general (even when those faces are seen unmasked). This conjecture is supported by findings that daily exposure to other-race faces reduces the other-race effect in older infants (Anzures et al., 2012), suggesting that even small, transient changes in face experience may affect face processing during a developmental sensitive period (Pascalis et al., 2020). In fact, recent work has shown that the processing of upright versus inverted masked faces (one hallmark of holistic face processing) changes with experience and/or development in infants (Galusca et al., 2023).

To test these predictions, we adapted the composite face method used by de Heering et al. (2007) to study the composite face effect in 4-, 5-, and 6-year-old children. Unlike de Heering et al., however, we conducted our study on an online platform rather than in a laboratory setting. Conducting such a study on an online platform has two principal advantages. First, it is easier to obtain a large sample size and, in the process, to better satisfy statistical assumptions. Second, the data can be collected from a more diverse participant sample than is usually possible in a typical laboratory study (Sheskin et al., 2020).

The current study consisted of two experiments. Experiment 1 was a conceptual replication of the de Heering et al. (2007) study in which we presented upright composite faces and investigated children's ability to discriminate them. Specifically, children saw a set of spatially aligned and spatially misaligned same and different top halves of faces combined with different bottom halves of faces in a two-alternative forced-choice task and were asked whether the tops of the faces were the same or different. In the case of holistic processing, people are typically poorer at discriminating the tops of faces in the *aligned–same* test trials than in the *misaligned–same* test trials. We examined whether 4- to 6-year-old children would show evidence of holistic processing, consistent with prior work showing holistic processing in early childhood (Crookes & McKone, 2009; de Heering et al., 2007; Macchi Cassia et al., 2009), or whether COVID-related masking may have affected the children's ability to process faces in a holistic fashion. As an exploratory follow-up, we also investigated whether the degree of visual disruption (i.e., exposure to masked faces) may relate to holistic face processing. To do so, we administered a questionnaire to the children's parents to measure their children's exposure to masked faces and examined the correlation between the degree of mask exposure and children's performance on the composite face task. This questionnaire was developed for the purposes of this study given that no existing questionnaires, to our knowledge, had been validated to measure experience with masked faces during the pandemic. However, because the reliability and validity of this questionnaire are unknown, we consider all analyses related to them as supplemental to our main analyses.

Experiment 2 was designed to complement Experiment 1 and provide convergent evidence of holistic face processing by testing children's task performance with inverted faces. Thus, in Experiment 2 we presented the same set of composite faces presented in Experiment 1 except that this time we disrupted holistic/configural processing by inverting the faces. If the children engaged in holistic face processing in Experiment 1, then we expected the face inversion in Experiment 2 to disrupt it and, thus, that the children would no longer exhibit poorer discrimination in *aligned–same* trials than in *misaligned–same* trials. One important design feature of Experiment 2 was that we retested a subsample of the same children who we tested in Experiment 1. This enabled us to control for individual differences when comparing the results from both experiments.

Experiment 1: Upright faces

We had two primary aims in Experiment 1: (a) replicate de Heering et al.'s (2007) composite face effect in 4- to 6-year-old children and (b) determine whether and to what extent face coverings of social partners and community members during the COVID-19 pandemic had a detrimental effect on young children's holistic face processing.

Method

Participants

We recruited and tested 142 4- to 6-year-old children on Lookit (https://lookit.mit.edu), an online recruitment and testing platform (Scott et al., 2017; Scott & Schulz, 2017) in August and September of 2021. Of the original sample, 8 children did not provide a complete data set either because they failed to complete the experiment or because technical problems prevented them from completing it, and 1 additional child was excluded for reporting an autism diagnosis. The remaining 133 children (62 female gender, 1 other gender) completed the experiment and, thus, provided usable data (M_{age} = 5.39 years, *SD* = 0.88, range = 4.03–6.99). This final sample of children was divided into separate age groups for analytic purposes and consisted of a group of 4-year-olds (*n* = 49; M_{age} = 4.44 years, *SD* = 0.28; 20 female), 5-year-olds (*n* = 47; M_{age} = 5.50 years, *SD* = 0.24; 23 female, 1 other gender), and 6-year-olds (*n* = 37; M_{age} = 6.53 years *SD* = 0.31; 19 female). For analyses on gender, we used a binary variable for male versus non-male (grouping together participants who identified as female and other gender).

We tested 41 children during a first phase of the experiment and 92 additional children during a second phase of the experiment. The two phases of the experiment were identical except for two minor changes instituted during the second phase. The first change was based on parent feedback and included a friendly task reminder at the start of every 10 test trials (i.e., a reminder that the children needed to respond whether the purple parts of the faces were the "same" or "different" by clicking on one of two buttons visible on the screen corresponding to these choices). The second change consisted of the addition of another question to the COVID-19 demographics questionnaire administered to the children's parents to help ascertain their children's exposure to masked faces (see below).

The parents of the majority of the children tested in this experiment identified as either White (52.63%), biracial (24.06%), or Asian (13.53%) and as living in either a suburban environment (57.89%) or an urban environment (38.35%). Overall, the average educational level of the children's parents was relatively high (bachelor's degree: 30.08%; graduate/professional degree: 56.39%), as was their wealth status (annual income ≥$100,000: 52.63%).

Apparatus and stimuli

We created composite face stimuli from high-resolution face images (Morrison et al., 2017) retrieved from the Open Science Framework (https://osf.io/g27wf). Faces were grayscale images of White males and females (19–30 years of age) looking directly into the camera with a neutral expression. We presented 24 pairs of composite face stimuli to each participant. Half of these pairs consisted of female composite face pairs, and the other half consisted of male composite face pairs. Given that misalignment of inner face features can reduce the composite face effect (Curby & Entenman, 2016; but see Kurbel et al., 2021, for robust results regardless of perceptual fit), we ensured as much as possible that each individual face was paired with a same-gender face of similar size/shape and skin tone. In addition, as recommended for the composite face task (Rossion & Retter, 2015), we included a small gap between the top and bottom halves of each composite face.

For each identity pair, we created 8 different composite faces from combinations of the top and bottom halves of the faces (these can be seen in Fig. 1A). Four of these composite faces consisted of spatially aligned top and bottom halves of faces, and the other four consisted of spatially misaligned top and bottom halves of faces (in the misaligned composite faces, the top half of the face was shifted ~1.2 cm to the left of the bottom half of the face). As can be seen in Fig. 1A, Composite Face AA consisted of the top and bottom halves of Identity A, Composite Face AB consisted of the top half of Identity A and the bottom half of Identity B, Composite Face BA consisted of the top half of Identity B and the bottom half of Identity A, and Composite Face BB consisted of the top and bottom halves of Iden-

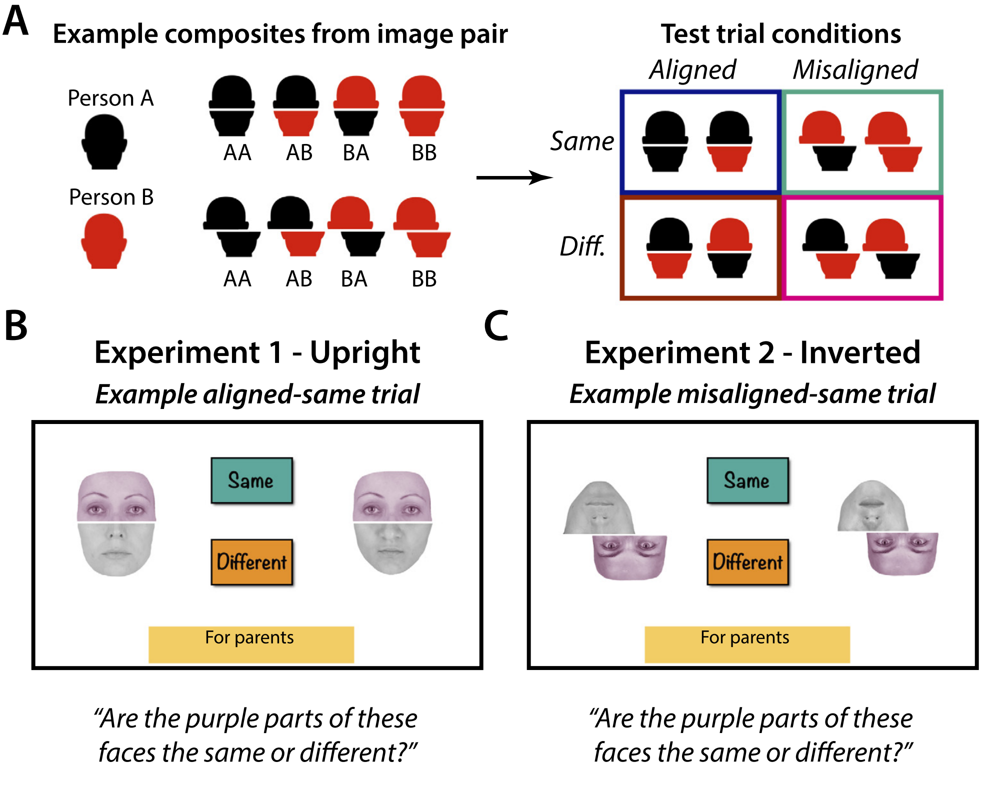

Fig. 1. Stimuli presented in Experiments 1 and 2. (A) Cartoonized example of how face identity pairs were combined to make different composite faces. On a given test trial, participants saw composite faces presented on the left and right sides of the screen according to 4 different conditions: *aligned-same*, *misaligned-same*, *aligned-different*, and *misaligned-different*. (B) Example of a stimulus trial in Experiment 1. (C) Example of a stimulus trial in Experiment 2.

tity B. To minimize the impact of external face features, we removed all hair and ears from the original images by using Adobe Photoshop 2020 and added a slight purple–pink tint to the top halves of the faces to draw children's attention to the top halves (de Heering et al., 2007). The full stimulus set is available at a public GitHub link (https://github.com/tristansyates/Lookit-Holistic-Face).

Procedure

Once parents logged on to the Lookit web page, they were asked to read a consent form and affirm their willingness to have their children participate in the study. Then, the children were asked to provide verbal assent after hearing a child-friendly version of the consent form. Finally, parents saw a set of written instructions informing them how to prepare their children for the experiment and were asked to refrain from helping their children in any way.

The first part of the experiment consisted of two practice trials. During the first of these trials, children saw a pair of spatially aligned composite faces where the top halves were of different faces. One composite face was presented on one side of the screen, and the other composite face was presented on the other side of the screen. Children were asked whether the "purple parts" (i.e., the tops) of these composite faces were the same or different. An incorrect response elicited a recorded message that asked them to try again, whereas a correct response elicited a recorded message that said, "Great job. The purple parts of these faces are different." During the second practice trial, children saw a pair of spatially misaligned composite faces where the top halves were of the same face and were once again asked whether the "purple parts" of these faces were the same or different. Again, an incorrect response elicited a recorded message asking them to try again, whereas a correct response elicited a recorded message that said, "Great job. The purple parts of these faces are the same." In each case, the stimulus pairs remained on the screen until children chose the correct answer.

As soon as the practice trials were completed, the children were given 36 test trials during which we presented four different types of stimulus pairs. These pairs were (a) *aligned–same,* where the top halves of the two composite faces depicted the same identity (e.g., AA and AB) and where the top and bottom halves were horizontally aligned, (b) *aligned–different,* where the top halves of two composite faces depicted different identities (e.g., BA and AB) and where the top and bottom halves were horizontally aligned, (c) *misaligned–same*, where the top halves of two composite faces depicted the same identity but were horizontally offset, and (d) *misaligned–different,* where the top halves of two composite faces depicted different identities but were horizontally offset (Fig. 1A, right). Fig. 1B shows an example of the types of stimuli and response buttons presented during an *aligned–same* trial.

Each child received a different random sequence of 36 test trials. Consistent with previous work (de Heering et al., 2007), we oversampled "same" trials under the assumption that in some of these trials the faces would be perceived as "different" if children performed holistic face processing. Of the 24 identity pairs, 6 were assigned to the *aligned–different* condition and 6 were assigned to the *misaligned–different* condition (for a total of 12 trials). The remaining 12 identity pairs were assigned to the *aligned–same* and *misaligned–same* conditions (12 trials each for a total of 24 trials). Specifically, for a pair of Identities A and B, participants would see either AB–AA in the aligned condition and BA–BB in the misaligned condition or vice versa. Note that this meant that the same bottom halves of faces would be repeated a second time throughout the task, but the top halves of faces were always novel identities across the trials. Trials in the *different* conditions were considered filler trials and were analyzed separately from the trials in the *same* condition.

During each trial, children were prompted to respond whether the faces were the same or different by clicking on one of two buttons visible on the screen corresponding to these choices (see Fig. 1B). Those children who participated during the second phase of this experiment were also reminded after 10, 20, and 30 trials that they were supposed to answer whether the "purple parts" of the face were the same or different. A click of one of the two response buttons was required to advance the experiment to the next trial. Children had unlimited time to respond and were allowed to let their parents click in their stead. Nonetheless, we restricted all analyses to those trials where the response time fell within 2 standard deviations of that child's average response time. Crucially, parents were permitted to click the response button only after their children first audibly stated a response to the question. We were able to confirm children's responses from the video-recordings of the test session where we could hear their verbal response and/or see their click.

Once children completed the experiment, parents were asked to fill out a COVID-19 demographics questionnaire (see Appendix in online supplementary material) and the experiment ended with a debriefing page. The purpose of the COVID-19 questionnaire was to quantify the degree to which children experienced partially visible faces in their daily life during both the initial phase of the COVID-19 pandemic (March 2020–March 2021) and after vaccines became more widely available in the United States (March 2021–September 2021). Parents reported how often children saw members of their household and members of their community wearing face masks (on a scale ranging from *daily* to *never*). All parents were prompted to answer whether their children attended in-person daycare/school that required interactions with masked individuals (*yes, no,* or *sometimes*) in the past and present, and a subset of parents gave the specific number of hours their children spent in daycare/school per week. Some participants answered *no* to the former question about daycare status and either left the numeric question blank or were not asked the question (as was the case for the children tested during the first phase). For these children, we coded the number of hours spent in daycare as 0. We also asked parents to report the state and the nature of the mask mandates in their area. Finally, parents indicated whether their children could tell people apart even if they were wearing masks or if their children sometimes had difficulty in telling masked people apart. In total, we obtained complete COVID-19 demographics information from 96 of the 133 children. All the procedures of this experiment were approved by the local institutional review board.

Results

First, we wanted to ensure that the task reminder introduced during the second phase of this experiment did not differentially affect responses. Therefore, we conducted separate analyses of the data from the two phases of testing. These analyses indicated that the main results were not affected by the addition of the task reminders (see Fig. S1 in supplementary material), and as a result we collapsed the data from the two phases of the experiment for all subsequent analyses.

Response accuracy

The data of primary interest were the accuracy scores obtained in the *same* trials. Specifically, lower accuracy scores in the *aligned–same* trials than in the *misaligned–same* trials is generally considered to reflect holistic processing. The data of secondary interest were the accuracy scores obtained in the different trials. In this case, the accuracy scores in the *aligned–different* trials and in the *misaligned–different* trials should be the same. These *different* trial scores indicate how well children were able to detect differences when the top halves of the composite faces actually differed and, thus, provide a baseline against which to evaluate the accuracy data from the *same* trials. To statistically assess response accuracy, we performed mixed repeated-measures analysis of variance (ANOVA) models with alignment as a within-participants factor, age and gender as between-participants factors, and participant as a random effect. For all analyses, we included gender as a predictor given prior research showing that there may be differences in holistic face processing even in early childhood (Stajduhar et al., 2022).

Same trials. Fig. 2 shows the accuracy scores in the *same* trials, and it can be seen that the predicted effect was present at each age. That is, at each age children exhibited lower accuracy scores when the same top halves of a face were aligned with the bottom halves of two different faces than when they were misaligned. A mixed repeated-measures ANOVA of the accuracy scores yielded significant main effects of alignment, $F(1, 127) = 70.57$, $p < .001$, $\eta^2_G = .104$, and age, $F(2, 127) = 6.86$, $p = .001$, $\eta^2_G = .079$, but no main effect of gender, $F(1, 127) = 0.99$, $p = .323$, $\eta^2_G = .006$. The ANOVA did not yield any significant interactions [Age × Alignment: $F(2, 127) = 1.34$, $p = .264$, $\eta^2_G = .004$; Gender × Alignment: $F(1, 127) = 0.31$, $p = .577$, $\eta^2_G = .0005$; Age × Gender: $F(2, 127) = 1.19$, $p = .307$, $\eta^2_G = .015$; Age × Gender × Alignment: $F(2, 127) = 0.868$, $p = .422$, $\eta^2_G = .003$].

Although the main effect of age is not informative with regard to the difference in accuracy in the critical *aligned–same* versus *misaligned–same* trials, it is nonetheless informative with regard to overall accuracy as a function of age. Therefore, given the significant age effect, we compared average response accuracy scores in the *same* trials across age with two-tailed, Bonferroni-corrected *t* tests.

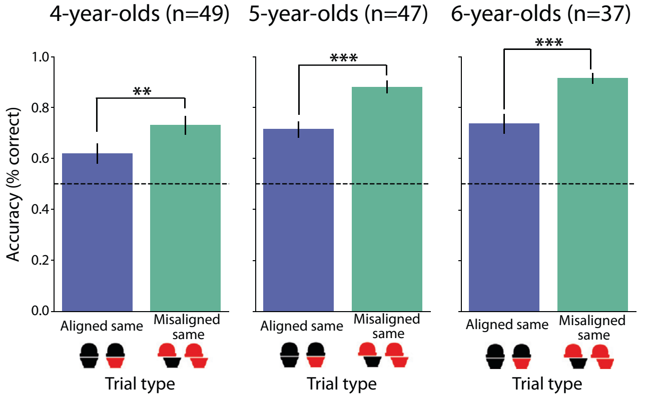

Fig. 2. Accuracy in the *aligned-same* and *misaligned-same* trials as a function of age in Experiment 1 (upright faces). Asterisks denote Bonferroni-corrected significant differences (*** $p < .001$, * $p < .05$) and error bars represent standard errors of the mean.

As expected, 4-year-olds were significantly less accurate (M = .68) than both 5-year-olds (M = .80), $t(90) = -2.84$, p = .017, $d = -0.579$, and 6-year-olds (M = .83), $t(83) = -3.48$, p = .002, $d = -0.739$, whereas 5-year-olds were not significantly less accurate than 6-year-olds, $t(80) = -0.76$, p = 1.00, $d = -0.166$. These results indicate that response accuracy improved from 4 to 5 years of age and then remained at the same level from 5 to 6 years of age.

Importantly, the absence of an Age × Alignment interaction indicates that the magnitude of the difference in accuracy scores across the *aligned–same* and *misaligned–same* trials did not differ across age. To determine whether this difference was statistically significant at each age, we performed two-tailed, Bonferroni-corrected paired t tests comparing accuracy on *aligned–same* and *misaligned–same* trials within each age group. As Fig. 2 shows, accuracy on *aligned-same* trials was significantly lower than accuracy on *misaligned–same* trials in 4-year-olds (*aligned–same* M = .62, misaligned–same M = .73), $t(48) = -3.15$, p = .008, $d = -0.450$, 5-year-olds (*aligned–same* M = .71, *misaligned–same* M = .88), $t(46) = -7.19$, $p < .001$, $d = -1.05$, and 6-year-olds (*aligned–same* M = .74, *misaligned–same* M = .92), $t(36) = -5.57$, $p < .001$, $d = -0.916$. Thus, children of all ages exhibited the expected effect of holistic processing.

Different trials. As indicated earlier, the *different* test trials provide an important check on the difference in accuracy scores obtained in the *same* test trials. A mixed repeated-measures ANOVA of the accuracy scores in the *different* trials (Fig. S2) yielded main effects of alignment, $F(1, 127) = 6.18$, p = .014, η^2_G = .012, and age, $F(2, 127) = 19.69$, $p < .001$, η^2_G = .190, but no effect of gender, $F(1, 127) = 0.88$, p = .349, η^2_G = .005, or any interactions [Age × Alignment: $F(2, 127) = 0.83$, p = .438, η^2_G = .003; Gender × Alignment: $F(1, 127) = 0.00$, p = .998, η^2_G = .00; Age × Gender: $F(2, 127) = 0.158$, p = .854, η^2_G = .002; Age × Gender × Alignment: $F(2, 127) = 0.73$, p = .486, η^2_G = .003].

To determine the source of the main effect of age, we compared the average accuracy scores across age with two-tailed, Bonferroni-corrected t tests. These indicated that 4-year-olds were significantly less accurate (M = .59) than 5-year-olds (M = .76), $t(94) = -3.27$, p = .004, $d = -0.668$, and 6-year-olds (M = .90), $t(65) = -7.83$, $p < .001$, $d = -1.62$, and that 5-year-olds were less accurate than 6-year-olds, $t(61) = -3.39$, p = .004, $d = -0.711$. These differences show that, as was the case for the *same* trials, the children's accuracy scores in the *different* trials improved, except that here they improved across all three ages.

To identify the source of the main effect of alignment, we compared the accuracy scores across the two alignment conditions with two-tailed, Bonferroni-corrected paired t tests within each age group. Even though accuracy was consistently greater in the *aligned–different* trials than in the *misaligned–different* trials, these differences were not statistically significant at any age after correction

[4-year-olds: *aligned–different* M = .63, *misaligned–different* M = .55), $t(48) = 1.73$, $p = .268$, $d = 0.248$; 5-year-olds: *aligned–different* M = .79, *misaligned–different* M = .73), $t(46) = 2.00$, $p = .156$, $d = 0.291$; 6-year-olds: *aligned–different* M = .91, *misaligned–different* M = .90), $t(36) = 0.33$, $p = 1.00$, $d = 0.054$]. Thus, in contrast to the differences in accuracy scores across the alignment conditions observed in the *same* trials, there were no such differences in the *different* trials. This suggests that the overall main effect of alignment reflects the greater statistical power of the aggregated data from all three age groups. Furthermore, the trend was in the opposite direction relative to the *same* trials; here, children were slightly more accurate on aligned trials than on misaligned trials.

Relationship between accuracy and COVID-19 variables

Finally, we explored whether environmental factors related to the COVID-19 pandemic may have influenced (a) accuracy on the critical *aligned–same* trials and (b) the difference in accuracy for *aligned–same* versus *misaligned–same* trials. To reiterate, our initial motivation for examining the relation between the various measures in our questionnaire and accuracy scores was the theoretically reasonable question of whether exposure to masked faces might have negative effects on the developmental emergence of holistic face processing. For this analysis, we first recoded all categorical variables as ordinal variables and dropped any "prefer not to answer" responses and nonresponses, resulting in values for 11 COVID-19 questions from 96 participants (Table S1). Before examining the relationship between COVID-19 questionnaire data and response accuracy on the holistic face task, we first assessed whether there were any differences in children's experiences based on age group. As shown in Table S1, only one of the COVID-19 variables was significantly related to age after correcting for multiple comparisons: daycare/school status early in the pandemic (March 2020–March 2021; $\chi^2(6) = 24.23$, $p = .006$, $d = 1.16$). Follow-up Bonferroni-corrected tests indicated that the 4-year-olds attended daycare less early in the pandemic than the 5-year-olds, $\chi^2(4) = 17.57$, $p = .004$, $d = 1.18$, and the 6-year-olds, $\chi^2(4) = 23.95$, $p < .001$, $d = 1.53$, but that the 5- and 6-year-olds attended daycare equally often, $\chi^2(4) = 8.44$, $p = .230$, $d = 0.80$.

We then performed an exploratory factor analysis on the COVID-19 questionnaire data as a data-driven dimensionality reduction step (Fabrigar et al., 1999) using the Python package *FactorAnalyzer* (https://factor-analyzer.readthedocs.io/en/latest/index.html). Results from Bartlett's test of sphericity revealed that the correlation matrix of COVID-19 variables (Table S2) was significantly different from the identity matrix, $\chi^2(950, 96) = 442.06$, $p < .001$, indicating that dimensionality reduction would be appropriate. Furthermore, the Kaiser–Meyer–Olkin test of sampling adequacy revealed a mediocre but acceptable value of .65. We used the minimal residual solution with a varimax rotation for our exploratory factor analysis. Following prior work (Kaiser, 1960), we retained factors that had an eigenvalue greater than 1, resulting in a four-factor solution that cumulatively explained 59.81% of the variance in the COVID-19 questionnaire data. Visual inspection of the factor loadings (Table S3) revealed that measures of current and past daycare exposure loaded heavily onto the first and second latent factors, respectively. The degree of exposure children had to members of their household in wearing masks, both early and later in the pandemic, loaded heavily onto the third latent factor, whereas measures of the severity of masking in the community early in the pandemic loaded heavily onto the fourth latent factor. We assessed whether children's age was related to factor values and found this to be the case for the third latent factor, $F(2, 93) = 5.01$, $p = .034$, $\eta^2_G = .097$, but no other factors were significant after correcting for multiple comparisons [Factor 1: $F(2, 93) = 1.05$, $p = 1.00$, $\eta^2_G = .022$; Factor 2: $F(2, 93) = 4.33$, $p = .064$, $\eta^2_G = .085$; Factor 4: $F(2, 93) = 0.24$, $p = 1.00$, $\eta^2_G = .005$]. In post hoc two-sample t tests, we found that the age effect on the third latent factor was driven by a difference between values in 5-year-olds ($M = -.44$) and 6-year-olds ($M = .26$), $t(50) = -2.58$, $p = .039$, $d = -0.652$.

With this factor analysis in hand, we ran generalized linear models using the ordinary least squares function from the Python package *statsmodels* (Seabold & Perktold, 2010). First, we ran a model with the four factors, age, and gender as predictors of response accuracy on the *aligned–same* trials. The adjusted R^2 revealed that this model explained only 2.95% of the variance in accuracy. Although age remained a significant predictor of accuracy, $b = 0.08$, $t(89) = 2.50$, $p = .014$ (Table S4), none of the other factors contributed significantly (all ps > .09). Next, we used the same predictors to instead model the difference in accuracy on *misaligned–same* and *aligned–same* trials. The logic was that this difference measure may better capture holistic face processing by accounting for task accuracy more generally. The results of this model are shown in Table S5. Neither age nor any of the latent factors from the COVID-19 questionnaire data predicted the difference in children's accuracy for *misaligned-same* minus *aligned–same* trial accuracy (all *p*s > .31). Additional models that included ethnicity (White vs. non-White) were qualitatively similar (Tables S6 and S7). Nonetheless, it may be the case that experience with face masks had a subtler influence on holistic face processing, perhaps in the form of response times. However, general linear models with the four COVID-19-related factors, age, and gender as predictors did not predict response time on the *aligned–same* trials (Table S8) or the difference in response time on *aligned–same* and *misaligned–same* trials (Table S9). Taken together, these analyses probing a possible association between COVID-19 variables and holistic processing yielded no evidence of any significant associations.

Discussion

The aim of Experiment 1 was to test young children's ability to perceive faces in a holistic manner in the context of the COVID-19 pandemic and, thus, to determine whether exposure to masked faces might have deleterious effects on this ability. The results from Experiment 1 provided evidence that online testing of young children's face discrimination abilities is possible and that it yields reliable findings that replicate previous findings obtained in a more controlled experimental setting (de Heering et al., 2007). We found that 4-, 5-, and 6-year-old children exhibited the composite face effect and, thus, provided evidence of holistic face processing. Furthermore, even though the magnitude of the composite face effect did not differ as a function of age, we found that 5- and 6-year-old children had higher accuracy scores overall compared with 4-year-olds. Finally, we found that exposure to masked faces did not appear to have negatively affected children's holistic face processing.

Experiment 2: Inverted faces

The method used in Experiment 1 to test for the presence or absence of the composite face effect in children is based on a method used in past adult and developmental studies. Nonetheless, to increase confidence in our findings, we conducted a second experiment in which we employed the same procedures and presented the identical stimuli except that this time the faces were inverted. Inversion keeps the relational features and pixel values the same while it reduces the tendency to perceive faces in a holistic manner (Rossion, 2013). If the children tested in Experiment 1 were indeed responding to the composite faces as unitary entities, they should not treat the inverted composite faces as unitary in the current experiment. To test this prediction, we tested children (including a subsample of the children who were initially tested in Experiment 1) with identical but inverted faces. By testing some of the same children, we were able to control for between-participants differences and rule out this specific factor as contributing to any differences in performance across the two experiments.

Method

Participants

We recontacted a subset of the participants (n = 85) from Experiment 1 to participate in a follow-up study that we conducted from December 2021 to February 2022. Of the 85 contacted participants, 35 of them (15 female gender; M_{age} = 5.59 years, *SD* = 0.88, range = 4.19–7.15) participated in Experiment 2. An additional 8 children attempted to complete the task but did not finish all the test trials; they were not included in the final sample. Because the size of this subsample was much smaller than that in Experiment 1, we collected additional data from a set of 95 new participants (48 female gender, 1 non-answer; M_{age} = 5.27 years, *SD* = 0.91, range = 4.03–6.99) to increase overall statistical power. We tested an additional 10 children but excluded their data because 7 of them did not finish all the test trials, 2 accidentally participated in the study twice, and 1 had a parent-reported autism diagnosis. Unless specified, we report the results from all 130 participants who participated in Experiment 2. A binomial test revealed no difference in the gender distribution between Experiments 1 and 2

(47.4% non-male in Experiment 1 vs. 47.7% non-male in Experiment 2, binomial $p = 1.00$). In addition, there was no difference in the children's average age between the two experiments (5.39 vs. 5.36 years), $t(261) = 0.034$, $p = .759$. The final sample consisted of 4-year-olds ($n = 57$; $M_{age} = 4.48$ years, $SD = 0.25$; 25 female), 5-year-olds ($n = 32$; $M_{age} = 5.46$ years, $SD = 0.28$; 17 female), and 6- and 7-year-olds ($n = 41$; $M_{age} = 6.50$ years, $SD = 0.31$; 19 female, 1 non-answer). As in Experiment 1, the parents tended to identify as White (59.38%), biracial (13.81%), or Asian (15.62%) from suburban (66.67%) or urban (22.92%) areas and with high levels of education (graduate/professional degree: 58.33%; bachelor's degree: 32.29%) and wealth (families with annual income ≥$100,000: 48.96%).

During data acquisition for the first 35 children in Experiment 2, a temporary error on the Lookit server prevented these participants from completing the experiment during their first attempt and forced them to restart it. This raised the possibility that these participants' responses may have partly reflected practice effects. Fortunately, this was unlikely because stimulus presentation was randomized in the different trial conditions for each participant and meant that the same test trials were not administered across the participants' multiple attempts to complete the experiment. As a result, we adopted a two-pronged approach to the data analysis. First, in the principal analysis, we included all participants who completed a full session of the experiment regardless of whether or not they had previously attempted to do the experiment. Second, we also explored the possibility of practice effects (Fig. S3) by (a) restricting our analyses to participants who completed the experiment in one session ($n = 26$ of 35 participants collected during this error) and (b) relating the number of trials completed on previous attempts to children's performance.

Apparatus and stimuli

The stimuli for Experiment 2 were identical to those presented in Experiment 1 except that the composite faces were rotated 180 degrees to create inverted composite face images (Fig. 1C).

Procedure

The procedure for Experiment 2 was identical to that used in Experiment 1.

Results

Response accuracy

Same trials. If children's lower performance on *aligned–same* trials in Experiment 1 was due to holistic face processing, inverting the stimuli should increase accuracy on these trials given that inversion is known to disrupt holistic processing. However, if other factors such as response demands, executive

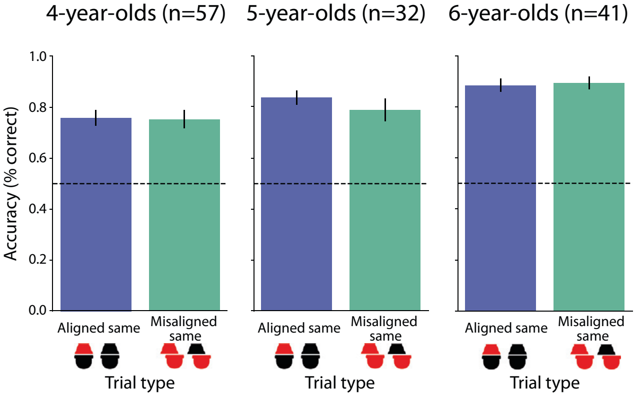

Fig. 3. Accuracy on the *aligned-same* and *misaligned-same* trials as a function of age in Experiment 2 (inverted faces). Error bars represent standard error of the mean.

control, and/or attention interfered with performance (Ventura et al., 2018), inverting the stimuli should have no effect on accuracy and we should find a similar pattern of findings as in Experiment 1.

As can be seen in Fig. 3, in contrast to Fig. 2, accuracy was not consistently lower in the *aligned–same* trials than in the *misaligned–same* trials. A mixed repeated-measures ANOVA on response accuracy yielded a main effect of age, $F(2, 124) = 5.85$, $p = .004$, $\eta^2_G = .069$, but no main effects of alignment, $F(1, 124) = 0.35$, $p = .553$, $\eta^2_G = .0006$, gender, $F(1, 124) = 0.13$, $p = .721$, $\eta^2_G = .0008$, or any interactions [Age × Alignment: $F(2, 124) = 0.68$, $p = .508$, $\eta^2_G = .002$; Gender × Alignment: $F(1, 124) = 0.21$, $p = .651$, $\eta^2_G = .0004$; Age × Gender: $F(2, 124) = 1.88$, $p = .157$, $\eta^2_G = .023$; Age × Gender × Alignment: $F(2, 124) = 0.18$, $p = .836$, $\eta^2_G = .0006$]. To determine the source of the main effect of age, we compared the average accuracy scores across age with two-tailed, Bonferroni-corrected *t* tests. These indicated that the 4-year-olds ($M = .75$) did not differ in accuracy from the 5-year-olds ($M = .81$), $t(78) = -1.35$, $p = .540$, $d = -0.289$, that the 5-year-olds did not differ in accuracy from the 6-year-olds ($M = .89$), $t(61) = -1.97$, $p = .161$, $d = -0.469$, but that the 4-year-olds were less accurate than the 6-year-olds, $t(96) = -3.56$, $p = .002$, $d = -0.706$. These results are consistent with our prediction that face inversion should interfere with the holistic face processing exhibited in Experiment 1. They also demonstrate that response accuracy increases with age regardless of face orientation.

Different trials. Figure S4 shows the results for the *different* trials. A mixed repeated-measures ANOVA of the response accuracy scores yielded a main effect of age, $F(2, 124) = 22.47$, $p < .001$, $\eta^2_G = .22$, but no main effect of alignment, $F(1, 124) = 0.02$, $p = .896$, $\eta^2_G = .00003$, or gender, $F(1, 124) = 0.21$, $p = .648$, $\eta^2_G = .001$, and it did not yield any interactions [Alignment × Age: $F(2, 124) = 2.04$, $p = .135$, $\eta^2_G = .007$; Gender × Alignment: $F(1, 124) = 0.66$, $p = .419$, $\eta^2_G = .001$; Age × Gender: $F(2, 124) = 0.58$, $p = .563$, $\eta^2_G = .007$; Age × Gender × Alignment: $F(2, 124) = 0.90$, $p = .410$, $\eta^2_G = .003$]. Follow-up two-sample, Bonferroni-corrected *t* tests of response accuracy across age indicated that the 4-year-olds were less accurate ($M = .48$) than both the 5-year-olds ($M = .73$), $t(70) = -4.35$, $p < .001$, $d = -0.945$, and the 6-year-olds ($M = .81$), $t(95) = -6.67$, $p < .001$, $d = -1.34$, but that the 5-year-olds did not differ from the 6-year-olds, $t(62) = -1.53$, $p = .396$, $d = -0.363$. Overall, this shows that the two older groups of children were more accurate than the youngest group of children in the inverted *different* trials.

Cohort effect. Finally, we ran ANOVAs with cohort as an additional between-participants factor to assess whether there were any differences in response accuracy between the 35 participants who had previously participated in Experiment 1 and the 95 participants who had not. We found no main effect of cohort on response accuracy, $F(1, 118) = 0.28$, $p = .597$, $\eta^2_G = .002$, and no interactions between cohort and other factors ($ps > .084$) for the *same* trials. Similarly, we found no main effect of cohort on response accuracy, $F(1, 118) = 0.15$, $p = .695$, $\eta^2_G = .001$, and no interactions between cohort and other factors for the *different* trials ($ps > .200$).

Comparison of Experiments 1 and 2

Finally, we compared the data from the 35 children who participated in both experiments to control for any individual differences that might have contributed to the differences between Experiments 1 and 2. Recall that the separate analyses of the accuracy scores from each experiment found no Age × Alignment interactions in the *same* trials (Figs. 2 and 3). Therefore, to compare the accuracy scores across the two experiments, we collapsed them across the three age groups. We expected that accuracy would differ in the upright versus inverted *aligned–same* trials, consistent with our findings of holistic processing in Experiment 1 but not Experiment 2. In other words, this analysis allowed us to directly ask whether holistic processing is specific to the upright condition (Experiment 1) and is no longer present when faces are presented upside down. For comparison, we also analyzed upright versus inverted *misaligned–same* trials, expecting to not see a difference in accuracy scores. Consistent with our prediction, paired *t* tests (uncorrected) revealed that accuracy was lower in the upright *aligned–same* trials ($M = .70$) than in the inverted *aligned–same* trials ($M = .82$), $t(34) = -2.64$, $p = .012$, $d = -0.490$, but that accuracy did not differ in the upright *misaligned–same* trials ($M = .85$) compared with the inverted *misaligned–same* trials ($M = .85$), $t(34) = -0.03$, $p = .976$, $d = -0.007$. Thus, overall, inversion had the predicted effect on accuracy in the aligned–same trials (Fig. 4).

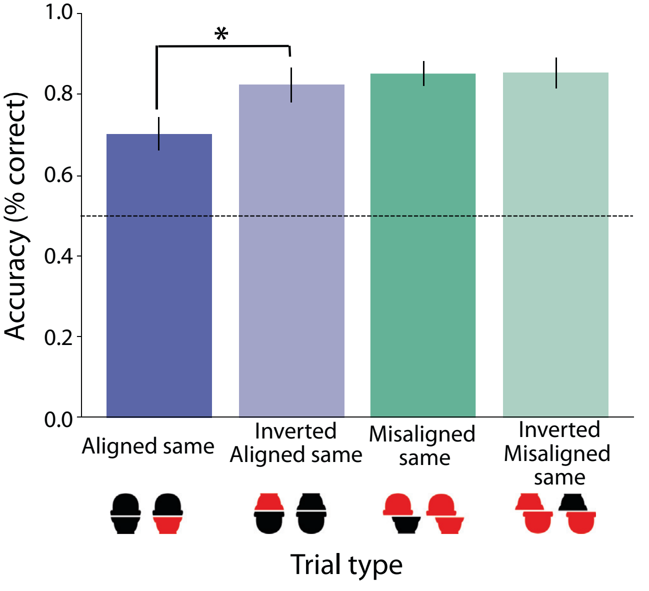

Fig. 4. Accuracy in the *aligned-same* and *misaligned-same* trials for upright vs. inverted composite faces in the children who participated in both Experiments 1 and 2. Error bars represent standard errors of the mean. ** $p < .01$.

Discussion

The results from Experiment 2 were consistent with our prediction that face inversion would disrupt holistic face processing. Children, including some of the same children who exhibited holistic face processing as evidenced by lower accuracy scores in the *aligned–same* trials than in the *misaligned–same* trials in Experiment 1, no longer exhibited this accuracy score difference when the faces were inverted.

General discussion

We investigated whether exposure to masked and, thus, partially visible faces during the COVID-19 pandemic may have had deleterious effects on the development of holistic face processing in early childhood. To do so, we used an online platform and measured 4-, 5-, and 6-year-old children's ability to process faces holistically along with their exposure to masked faces by administering a questionnaire to their parents. In Experiment 1, we presented pairs of composite faces composed of top halves that were either the same or different and bottom halves that were different. Results indicated that all age groups exhibited evidence of holistic face processing in that children exhibited lower accuracy scores in *aligned–same* trials than in *misaligned–same* trials. In Experiment 2, we presented the same face stimuli as in Experiment 1 except that this time we presented them in a spatially inverted position. As predicted, children no longer exhibited evidence of holistic face processing in that they did not exhibit lower accuracy scores in the *aligned–same* test trials than in the *misaligned–same* test trials. Finally, to investigate whether exposure to masked faces might have negatively affected holistic processing, we examined the relationship between accuracy scores in Experiment 1 and various measures of exposure. The results of this analysis showed that accuracy scores were not correlated with degree of mask exposure.

Our findings that 4-, 5-, and 6-year-old children exhibited holistic face processing replicate de Heering et al.'s (2007) findings of holistic face processing in the same age groups. Crucially, our replication was successful despite the fact that we tested children on an online platform rather than in a more controlled laboratory environment. This is a testament to the robust nature of holistic face processing at the ages tested here and demonstrates that the composite face effect is consistent enough that it can be obtained "in the wild" (i.e., in a child's home environment). Interestingly, and in contrast to de Heering et al. (2007), we observed an age-related increase in response accuracy between 4 and 5 to 6 years of age. This developmental improvement is in line with findings that children's general perception of faces improves in childhood (Mondloch et al., 2002) and with recent work that children's holistic face processing improves with age (Ventura et al., 2018).

Even though we replicated prior findings using the partial design of the composite face task, our results differ from Ventura et al.'s (2018) findings that 4-year-olds do *not* exhibit robust holistic face processing when a complete design is used. One possible reason that might account for the difference between our 4-year-old results and those of Ventura et al. may be that, as argued by some (Murphy et al., 2017; Richler & Gauthier, 2014), the complete design is a more accurate measure of holistic face processing. Another possible reason may be that, given that working memory improves with age in early childhood (Gathercole et al., 2004), the inclusion of an additional working memory component in the Ventura et al. (2018) study may have actually hindered the youngest children's task performance. Finally, it should be noted that we presented adult faces, whereas Ventura et al. presented 8-year-old faces. This difference in type of face presented makes it possible that young children's holistic face processing is affected differentially by adult versus child faces. In sum, even though the difference in the 4-year-old findings is interesting, it should be noted that the primary purpose of this study was not to determine precisely *when* in development holistic face processing emerges but rather to determine whether everyday experience contributes to holistic face processing in early childhood. Future studies should investigate the possible role that working memory demands and the specific age of the test faces may play in holistic face processing in early childhood.

The current study extends prior studies in two important ways. First, we included the all-important inverted face condition in Experiment 2. Inversion is known to disrupt holistic face processing in children (Carey & Diamond, 1977), and, as expected, we found that the same children who exhibited holistic face processing in Experiment 1 no longer did so when the faces were inverted in Experiment 2. This indicates unequivocally that the results from Experiment 1 reflect holistic processing. Second, we investigated the possible effects of altered visual experience with faces on holistic processing by measuring the degree of exposure to masked faces during the COVID-19 pandemic and by examining the statistical relationship between exposure and response accuracy in the composite face task. Unfortunately, we were unable to make *a priori* predictions regarding this correlation simply because we did not have any independent ways of determining what might constitute sufficient visual disruption of fully visible faces to have some measurable effect. Furthermore, we captured a rather coarse measure of children's experience with faces based on parental report by using a questionnaire that was not tested for reliability or validity. Finally, the one other measure that might have been informative, namely response times, did not yield any correlations with face experience (this was true even when analyses of response times were restricted to within 2 standard deviations of the mean). It should be noted, however, that response times were noisy because of the self-paced nature of the experiment and because of possible distractions in the home environment. Thus, the absence of a correlation between exposure and response accuracy is not surprising.

Future work will need to better quantify the extent and timing of children's face experience. One way to do so is to use head-mounted cameras on children while they navigate their world (Sullivan et al., 2022). Another way is to employ targeted measures such as gaze/selective attention (Bombari et al., 2009) and/or neural markers of face processing. Neural markers may be an especially effective way to examine the effects of face experience given evidence of neural differences in infants' processing of face identities pre- and post-lockdown (Yates, Ellis, & Turk-Browne, 2023). In the meantime, in the absence of a precise measure of face exposure in the current study, we can only speculate about the reasons why we found no relationship between degree of exposure to masked faces and accuracy scores.

One possible reason for the failure to obtain a correlation between face exposure and accuracy scores may be that holistic face perception is so robust by 4 years of age that it is no longer vulnerable to disruption by the relatively short period of exposure to partially visible faces. Another reason may be that exposure to fully visible faces at home was sufficient to overcome disrupted face experiences outside the home. A third reason may be that exposure to partially visible faces in early childhood has relatively subtle effects on the efficiency of face processing (Carnevali et al., 2022). Finally, it is possible that our results are restricted to children's processing of static images of faces rather than to the sorts of faces that children are usually exposed to in their everyday environment, namely dynamic faces usually seen in different poses and from different angles. In other words, we cannot rule out the possibility that mask wearing might have a more measurable effect on face processing in response to dynamic faces. This possibility is supported by findings that 10- to 12-month-old infants respond differently to the faces of other races when tested with static faces as opposed to dynamic faces (Minar & Lewkowicz, 2018). Accordingly, it is likely that the rather massive exposure that young children have acquired over the first years of their life to faces in many different poses, seen at different angles, and seen producing myriad expressions reinforces their ability to detect subtle differences. Presumably, this enables young children to detect differences even when they are inherent in static faces and this, in turn, enables them to process even static faces in a holistic manner (Favelle & Palmisano, 2012; Gray et al., 2017; McKone, 2008).

Although our questionnaire measured variables that reflected children's exposure to partially visible faces, there are additional variables that might have been relevant for us to consider when exploring a possible relationship between face masking and face perception. For instance, probable mental health problems, including feelings of loneliness and fear of illness, increased during the pandemic in children (Newlove-Delgado et al., 2021). Given that negative emotional states can decrease holistic face processing (Curby et al., 2012), it could be that we might have found a relationship between masking and holistic face processing if we had obtained measures of stress and anxiety. This may be especially true for children who experienced the illness or death of a family member. In addition, it could be that measures of maternal anxiety, which skyrocketed during the pandemic (Hessami et al., 2022), might have revealed effects of masking on holistic processing via effects on children's general socioemotional and cognitive well-being (Rogers et al., 2020).

Even though the current study did not uncover a relationship between exposure to masked faces and holistic face processing, one of its unique and valuable features is that it demonstrated that it is possible to study the development of face perception, in general, and the development of the composite face effect, in particular, outside of the traditional laboratory. To our knowledge, only one other study to date has tested children's face recognition abilities by using an online platform (Stajduhar et al., 2022). There are two notable advantages to using an online platform. First, data can be obtained from many children over a shorter period of time than in a laboratory-based study. Second, ideally speaking, an online platform offers the possibility of reaching more diverse populations than those that often participate in typical lab-based studies. Unfortunately, in the current study, we were not able to capture a population of children who were ethnically or socioeconomically more diverse than the populations that participate in typical laboratory studies. A likely reason for this is that reaching more diverse populations may be limited by the differential access that such populations may have to the technology required to participate in an online study. Therefore, our results should be interpreted as reflecting a particular demographic. Given this, it could be that face processing in certain groups of children may be more or less affected by masking during the COVID-19 pandemic, perhaps because of differences in the number and types of faces that they see (Sangrigoli et al., 2005). Moreover, like many other studies of face processing, we presented White (adult) faces. Such faces might be appropriate for White children, but not for children from other races or ethnicities. Although we found no evidence that children's own race/ethnicity affected holistic face processing on this task (Tables S6 and S7), we were unable to test whether children's experiences with different races or ethnicities may have influenced holistic face processing. In addition, because children presumably saw their adult caregivers without a mask, it may be the case that holistic processing is not disrupted for adult faces but perhaps is disrupted for other faces that are more likely to be seen masked (e.g., other children, older adults). If holistic face processing depends, in part, on the specific early experience that children have with faces of a specific race or set of races (or a specific age or set of ages; Rhodes & Anastasi, 2012), then it is highly likely that holistic face processing may be most robust for the statistically most frequent face category in a child's everyday life.

In conclusion, our findings do not permit us to distinguish between the possibility that young children's holistic face processing may be too robust by 4 years of age to resist relatively short-term disruption in viewing fully visible faces and the possibility that our exposure measures may have missed variables that might have revealed a relationship between masked face experience and holistic face processing. This, in turn, means that the current results cannot provide answers to questions about the length of the sensitive period for the developmental emergence of holistic face processing. Nonetheless, our findings make two important contributions. First, they add to the growing body of evidence indicating that holistic face processing is sufficiently advanced as early as 4 years of age. Second, they demonstrate that it is possible to obtain evidence of holistic processing in young children even under less than ideal conditions by testing them in their everyday ecological setting. The latter finding opens up lots of new opportunities for exploring other aspects of face perception in early childhood.

Data availability

The data and code that support the findings of this study are openly available on GitHub (https://github.com/tristansyates/Lookit-Holistic-Face).

Acknowledgments

This research was supported in part by a grant from the National Science Foundation (BCS 1749507) awarded to D.J.L. We thank all the families who participated and the Lookit research team members for technical support and the use of their experiment platform. Finally, we thank the members of the Language Learning and Multisensory Brain (LLAMB) Lab at Haskins Laboratories for their helpful advice on the use of Lookit.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2023.105676.

References

Anzures, G., Wheeler, A., Quinn, P. C., Pascalis, O., Slater, A. M., Heron-Delaney, M., Tanaka, J. W., & Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology, 112*(4), 484–495.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science, 17*(2), 159–163.

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception, 38*(10), 1508–1521.

Boutet, I., Nelson, E. A., Watier, N., Cousineau, D., Béland, S., & Collin, C. A. (2021). Different measures of holistic face processing tap into distinct but partially overlapping mechanisms. *Attention, Perception, & Psychophysics, 83*, 2905–2923.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*(4275), 312–314.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1*(2–3), 253–274.

Carnevali, L., Gui, A., Jones, E. J., & Farroni, T. (2022). Face processing in early development: A systematic review of behavioral studies and considerations in times of COVID-19 pandemic. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.778247. Article 778247.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111*(2), 219–247. https://doi.org/10.1016/j.cognition.2009.02.004.

Curby, K. M., & Entenman, R. (2016). Framing faces: Frame alignment impacts holistic face perception. *Attention, Perception, & Psychophysics, 78*(8), 2569–2578. https://doi.org/10.3758/s13414-016-1194-4.

Curby, K. M., Johnson, K. J., & Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition & Emotion, 26*(1), 93–102.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96*(1), 57–70. https://doi.org/10.1016/j.jecp.2006.07.001.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3). https://doi.org/10.1037/1082-989X.4.3.272. Article 272.

Favelle, S. K., & Palmisano, S. (2012). The face inversion effect following pitch and yaw rotations: Investigating the boundaries of holistic processing. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00563. Article 563.

Galusca, C. I., Clerc, O., Chevallier, M., Bertrand, C., Audeou, F., Pascalis, O., & Fort, M. (2023). The effect of masks on the visual preference for faces in the first year of life. *Infancy, 28*(1), 92–105.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190.
Gray, K. L., Murphy, J., Marsh, J. E., & Cook, R. (2017). Modulation of the composite face effect by unintended emotion cues. *Royal Society Open Science, 4*(4). https://doi.org/10.1098/rsos.160867. Article 160867.
Hessami, K., Romanelli, C., Chiurazzi, M., & Cozzolino, M. (2022). COVID-19 pandemic and maternal mental health: A systematic review and meta-analysis. *Journal of Maternal-Fetal & Neonatal Medicine, 35*(20), 4014–4021.
Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141–151.
Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084–1089.
Kurbel, D., Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2021). The composite face effect is robust against perceptual misfit. *Attention, Perception & Psychophysics, 83*(6), 2599–2612. https://doi.org/10.3758/s13414-021-02279-0.
Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15*(11), 762–768.
Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*(2), 292–315.
Macchi Cassia, V., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science, 12*(2), 236–248.
Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379–383.
Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.
Maurer, D., Lewis, T. L., & Mondloch, C. J. (2005). Missing sights: Consequences for visual cognitive development. *Trends in Cognitive Sciences, 9*(3), 144–151.
Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56*(2), 154–178.
McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 310–327.
Minar, N. J., & Lewkowicz, D. J. (2018). Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech. *Developmental Science, 21*(4). https://doi.org/10.1111/desc.12604. Article e12604.
Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*(5), 553–566.
Mondloch, C., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite face effect in six-year-old children: Evidence of adult-like holistic face processing. *Visual Cognition, 15*(5), 564–577.
Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception. *PLoS One, 12*(9). https://doi.org/10.1371/journal.pone.0185093. Article e185093.
Murphy, J., Gray, K. L., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review, 24*(2), 245–261.
Newlove-Delgado, T., McManus, S., Sadler, K., Thandi, S., Vizard, T., Cartwright, C., & Ford, T. (2021). Child mental health in England before and during the COVID-19 lockdown. *The Lancet Psychiatry, 8*(5), 353–354.
Noyes, E., Davis, J. P., Petrov, N., Gray, K. L., & Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science, 8*(3). https://doi.org/10.1098/rsos.201169. Article 201169.
Pascalis, O., de Martin de Viviés, X., Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 666–675.
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.
Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*(5571), 1321–1323.
Quinn, P. C., Uttley, L., Lee, K., Gibson, A., Smith, M., Slater, A. M., & Pascalis, O. (2008). Infant preference for female faces occurs for same- but not other-race faces. *Journal of Neuropsychology, 2*, 15–26.
Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146–174.
Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin, 140*(5), 1281–1302.
Rogers, A., Obst, S., Teague, S. J., Rossen, L., Spry, E. A., Macdonald, J. A., Sunderland, M., Olsson, C. A., Youssef, G., & Hutchinson, D. (2020). Association between maternal perinatal depression and anxiety and child and adolescent development: A meta-analysis. *JAMA Pediatrics, 174*(11), 1082–1092.
Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition, 21*(2), 139–253.
Rossion, B., & Retter, T. L. (2015). Holistic face perception: Mind the gap! *Visual Cognition, 23*(3), 379–398.
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.
Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind, 1*(1), 15–29.
Scott, K., & Schulz, L. (2017). Lookit (Part 1): A new online platform for developmental research. *Open Mind, 1*(1), 4–14.
Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In S. van der Walt & J. Millman (Eds.), *9th Python in Science Conference (SciPy 2010).* https://doi.org/10.25080/Majora-92bf1922-011.
Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences, 24*(9), 675–678.

Simion, F., & Di Giorgio, E. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00969. Article 969.
Simion, F., Leo, I., Turati, C., Valenza, E., & Dalla Barba, B. (2007). How face specialization emerges in the first months of life. *Progress in Brain Research, 164*, 169–185.
Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umiltà, C. (2002). Newborns' preference for up–down asymmetrical configurations. *Developmental Science, 5*(4), 427–434.
Stajduhar, A., Ganel, T., Avidan, G., Rosenbaum, R. S., & Freud, E. (2022). Face masks disrupt holistic processing and face perception in school-age children. *Cognitive Research: Principles and Implications, 7*(1). https://doi.org/10.1186/s41235-022-00360-2. Article 9.
Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2022). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind, 5*, 20–29.
Turati, C., Di Giorgio, E., Bardi, L., & Simion, F. (2010). Holistic face processing in newborns, 3-month-old infants, and adults: Evidence from the composite face effect. *Child Development, 81*(6), 1894–1905.
Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: What is crucial? *Developmental Psychology, 38*(6), 875–882.
Turati, C., Valenza, E., Leo, I., & Simion, F. (2005). Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. *Journal of Experimental Child Psychology, 90*(3), 255–273.
Ventura, P., Leite, I., & Fernandes, T. (2018). The development of holistic face processing: An evaluation with the complete design of the composite task. *Acta Psychologica, 191*, 32–41.
Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*(3), 437–456.
Wang, Z., Ni, H., Zhou, X., Yang, X., Zheng, Z., Sun, Y. H. P., Zhang, X., & Jin, H. (2023). Looking at the upper facial half enlarges the range of holistic face processing. *Scientific Reports, 13*(1). https://doi.org/10.1038/s41598-023-29583-z. Article 2419.
Wang, Z., Quinn, P. C., Jin, H., Sun, Y. H. P., Tanaka, J. W., Pascalis, O., & Lee, K. (2019). A regional composite-face effect for species-specific recognition: Upper and lower halves play different roles in holistic processing of monkey faces. *Vision Research, 157*, 89–96.
Yates, T. S., Ellis, C. T., & Turk-Browne, N. B. (2023). Face processing in the infant brain after pandemic lockdown. *Developmental Psychobiology, 65*. https://doi.org/10.1002/dev.22346. Article e22346.