

Robust Holistic Face Processing in Early Childhood During the COVID-19 Pandemic

Highlights

- We report an online replication of the composite face effect in children 4- to 6-years of age tested during the COVID-19 pandemic
- Face inversion disrupted the composite face effect, indicating children process faces holistically
- Children's performance was not related to latent variables from a parent questionnaire on masking experience during the pandemic
- Results indicate that the pandemic may have had a minimal impact on children's face processing

Abstract

The timing of the developmental emergence of holistic face processing and the role that experience plays in it are somewhat controversial topics. Using an online testing platform, we investigated holistic face perception in 4-6-year-old children with a two-alternative forced-choice task in which they saw pairs of composite faces and had to decide whether the faces were the same or different. We also assessed children's exposure to masked faces during the COVID-19 pandemic with a parental questionnaire to determine whether experience with masked faces may have negatively affected holistic processing. We found that children of all ages performed holistic face processing when the faces were upright (Experiment 1) but not when they were inverted (Experiment 2), that response accuracy increased with age, and that there was no relationship between degree of exposure to masked faces and response accuracy. 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.

Keywords: early childhood, composite face effect, face recognition, experience, perception

Robust Holistic Face Processing in Early Childhood During the COVID-19 Pandemic

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, Le Grand, & Mondloch, 2002). Mature face processing develops gradually out of some initial perceptual biases that can be observed right at birth (Pascalis et al., 2011; Simion, Leo, Turati, Valenza, & Dalla Barba, 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, while the second is a bias for more elements in the upper rather than 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, Simion, Milani, & Umiltà, 2002), for non-face-like stimuli that exhibit these structural properties (Simion, Valenza, Cassia, Turati, & Umiltà, 2002), and for faces themselves (Cassia, Turati, & Simion, 2004).

These initial preferences observed at birth provide a foundation for the gradual emergence of face-specific responsiveness which, to a large extent, is driven by infants' everyday experiences (Pascalis, Fort, & Quinn, 2020). Infants begin to prefer faces over scrambled face configurations containing more elements in the upper part by 3 months of age (Turati, Valenza, Leo, & Simion, 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 nor discriminate own-vs. other-race (or species) faces, but do just a

few months later (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2007; Pascalis, Haan, & Nelson, 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, while 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 also operate in 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, Pallier, Argenti, Ventureyra, & de Schonen, 2005). This demonstrates that the plasticity initially observed in infancy that enables infants to incorporate the statistics of faces in their everyday environment continues into early childhood (Maurer, Lewis, & Mondloch, 2005). This, along with findings that face processing expertise continues to grow well into adolescence (Mondloch, Le Grand, & Maurer, 2002), makes it theoretically reasonable to hypothesize that everyday experience may have an effect on mature, holistic face processing in early childhood. Yet, no studies to our knowledge have investigated whether everyday experience in visually-typical children influences the development of holistic face processing.

Overall, findings on the developmental emergence of holistic face processing have provided a rather mixed picture. On the one hand, some studies have reported evidence of holistic face processing as early as 3 months of age (Turati, Di Giorgio, Bardi, & Simion, 2010),

and that a lack of patterned visual input early in life can have a lasting impact on holistic face processing into adulthood (Le Grand, Mondloch, Maurer, & Brent, 2004). On the other hand, some studies have reported that holistic face processing emerges between 3 and 4 years of age (Cassia, Picozzi, Kuefner, Bricolo, & Turati, 2009; Crookes & McKone, 2009; de Heering, Houthuys, & Rossion, 2007) or that children do not exhibit holistic face processing until 6 (Carey & Diamond, 1994; Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007; Ventura, Leite, & Fernandes, 2018) or even as late as 10 years of age (Mondloch et al., 2002). At this point, it is unclear why these estimates are so divergent. One possible explanation may be the specific ways in which composite face methods have been used to test for holistic face processing (Ventura et al., 2018). Crucially, however, and regardless of the ultimate reasons for disparities in the specific ages when holistic face processing emerges, it should be noted that asking whether experience affects holistic face processing only requires comparing across studies with similar designs.

The aim of the current study was to determine whether everyday experience in early childhood might affect holistic face processing. We hypothesized that exposure to partially-visible faces may have detrimental effects on holistic face perception in young children if such processing depends on exposure to fully-visible faces. To test this hypothesis, 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. Even though children continued to see fully-visible faces of family members during the initial lockdowns, once lockdowns were lifted and they returned to day care, pre-school, and/or kindergarten, children were exposed mostly to the top halves of other people’s faces.

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, Davis, Petrov, Gray, & Ritchie, 2021) while another found that children's holistic face processing is altered (even more so than in adults) when viewing masked individuals (Stajduhar, Ganel, Avidan, Rosenbaum, & Freud, 2022). These findings suggest that children who are deprived of fully-visible faces for considerable parts of their day may find it difficult to discriminate faces (even when seen unmasked). To test these predictions, we adapted the 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. (2007), however, we conducted our study on an online platform rather than in a laboratory setting. Experiment 1 was a conceptual replication of the de Heering et al. (2007) study with upright composite faces, while Experiment 2 was a replication of Experiment 1 except with faces presented in an inverted position.

In Experiment 1, 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 2-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* than in the *misaligned-same* test trials. We predicted that children would not exhibit evidence of holistic processing if COVID-related masking was sufficient to deprive them of the expected experience that may be required to respond to faces in a holistic fashion. We also investigated whether the degree of visual deprivation (i.e., exposure to masked faces) related 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.

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 the configural cues by inverting the face. Importantly, to control for individual differences, we re-tested a subsample of the same children that we tested in Experiment 1. 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 they would no longer exhibit poorer discrimination in *aligned-same* than in *misaligned-same* trials.

Experiment 1: Upright Faces

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

Method

Participants. We recruited and tested one hundred-and-forty-two 4-6 year-old children on Lookit (<https://lookit.mit.edu/>), an online recruitment and testing platform (Scott, Chu, & Schulz, 2017; Scott & Schulz, 2017) between August and September of 2021. Nine 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. The remaining 133 children (N = 62 female gender, N = 1 other gender) completed the experiment and, thus, provided usable data (Mean age = 5.39, SD = 0.88; range 4.03 and 6.99 years). This final sample of children was

divided into separate age groups for analytic purposes and consisted of a group of 4-year-olds ($N = 49$; Mean age = 4.44 SD = 0.28; 20 females), 5-year-olds ($N = 47$; Mean age = 5.50, SD = 0.24; 23 females, 1 other gender), and 6-year-olds ($N = 37$; Mean age = 6.53, SD = 0.31; 19 females). For analyses on gender, we used a binary variable for male vs. 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 majority of the children tested in this experiment identified as either White (52.63%), Biracial (24.06%), or Asian (13.53%) and as living either in a suburban (57.89%) or an urban (38.35%) environment. 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 equal to or greater than \$100,000: 52.63%).

Apparatus & Stimuli. We created composite face stimuli from high-resolution face images (Morrison, Wang, Hahn, Jones, & DeBruine, 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 12 female composite face pairs while the other half consisted of 12 male composite face pairs. Given that misalignment of inner face features can reduce the composite face effect (Curby & Entenman, 2016) (although see (Kurbel, Meinhardt-Injac, Persike, & Meinhardt, 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. Also, 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 Figure 1A). Four of these composite faces consisted of spatially-aligned top and bottom halves of faces while the other four of these faces consisted of spatially-misaligned top and bottom halves of faces (in the misaligned composite faces, the top half of the face was shifted approximately 1.2 cm to the left of the bottom half of the face). As can be seen in Figure 1A, composite face AA consisted of the top and bottom half 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 half of identity 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 we 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](#).

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 child participate in the study. Then, the

child was 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 child for the experiment and were asked to refrain from helping their child 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 while 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 to 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, each child was given 36 test trials during which we presented 4 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 (Figure 1A, right). Figure 1B shows an example of the sorts 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 in the second phase of this experiment also were 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. Crucially, however, parents were only permitted to click the response button after their child 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 either 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) 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 child 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 child spent in daycare/school a week. Some participants answered “no” to the former question about daycare status and either left the numeric question blank or were not asked it (as was the case for the children tested during phase 1). 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 child could tell people apart even if they are wearing masks, or if they sometimes had difficulty telling masked people apart. In total, we obtained complete COVID-19 demographics information from 96 of the 133 children. All of the procedures of this experiment were approved by the local institutional review board.

Results

First, we wanted to ensure that the task reminder introduced in 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 (Figure S1) and, as a result, we collapsed the data from the two phases of the experiment for all subsequent analyses.

Response Time

Response time can reflect task difficulty, with more difficult discriminations requiring greater processing time prior to a decision. Therefore, we examined response time to determine whether discrimination may have differed in degree of difficulty. We defined response time as the time interval between the end of the audible instruction to respond and the click of the mouse on one of the response buttons. Overall, the average response time was 4.44 s (median = 3.71 s; range: 0.42 to 24.48 s). Importantly, it should be noted that this average response time is a combination of the time it took children to respond after the audible instruction ended (approximately 3 s) and likely practice effects (i.e., it is possible that children learned to respond increasingly faster as trials progressed). As a result, it is likely that the most rapid response times reflect anticipation to respond after the audible instruction.

Same trials

To statistically assess response times on the *same* trials, we performed a mixed repeated-measures analysis of variance (ANOVA) on response times, with alignment as a within-subjects factor, age and gender as between-subjects factors, and participant as a random effect. The ANOVA indicated that response time was not affected by alignment ($F(1,127) = 0.33$, $p = 0.569$), age ($F(2,127) = 1.36$, $p = 0.260$), gender ($F(1,127) = 0.87$, $p = 0.352$), nor by any

combination of these factors (age x alignment: $F(2,127) = 1.80$, $p = 0.169$; gender x alignment: $F(1,127) = 3.28$, $p = 0.072$; age x gender: $F(2,127) = 0.15$, $p = 0.861$; age x gender x alignment: $F(2,127) = 0.11$, $p = 0.894$). These results indicate that task difficulty may have been the same regardless of face alignment.

Different trials

Using the same analytic approach as for the *same* trials, we examined response times during the *different* trials. The ANOVA indicated response time was not affected by alignment ($F(1,127) = 0.52$, $p = 0.472$), nor gender ($F(1,127) = 0.13$, $p = 0.720$), but that it differed as a function of age ($F(2,127) = 4.45$, $p = 0.014$). None of the interactions were significant (age x alignment: $F(2,127) = 1.10$, $p = 0.335$; gender x alignment: $F(1,127) = 0.05$, $p = 0.817$; age x gender: $F(2,127) = 0.36$, $p = 0.700$; age x gender x alignment: $F(2,127) = 0.76$, $p = 0.468$).

Two-sample *t*-tests that included a Bonferroni correction ($p = 0.017$), comparing response time on *different* trials across age, indicated that the age effect was due to 4-year-olds responding more slowly ($M = 5.38$) than 6-year-olds ($M = 3.63$; $t(78) = 2.82$, $p = 0.018$).

Response Accuracy

The data of primary interest were the accuracy scores obtained in the *same* trials. These scores reflect holistic processing. The data of secondary interest were the accuracy scores obtained in the *different* trial. These scores provide a baseline against which to evaluate the accuracy data from the *same* trials. That is, the accuracy scores from the *different* test trials indicate how well children were able to detect differences when the top halves of the composite faces actually differed.

Same trials

If children engaged in holistic processing in a manner similar to that of adults, their accuracy scores in the critical *aligned-same* trials should have been lower than in the *misaligned-same* trials. Figure 2 shows the accuracy scores and, as can be seen, the predicted effect was present at each age. That is, at each age, children exhibited lower accuracy 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 ($F(1,127) = 67.94, p < 0.001$) and age ($F(2,127) = 7.26, p = 0.001$) but no main effect of gender ($F(1,127) = 0.68, p = 0.410$). The ANOVA did not yield any significant interactions (age x alignment interaction: $F(2,127) = 1.47, p = 0.234$; gender x alignment: $F(1,127) = 0.35, p = 0.557$; age x gender: $F(2,127) = 1.46, p = 0.236$; age x gender x alignment: $F(2,127) = 1.26, p = 0.288$).

Although the main effect of age is not informative with regard to the difference in accuracy in the critical *aligned-same* vs. *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. As expected, 4-year-olds were significantly less accurate ($M = 0.67$) than both 5-year-olds ($M = 0.79; t(90) = -2.79, p = 0.006$) and 6-year-olds ($M = 0.83; t(83) = -3.67, p < 0.001$) whereas 5-year-olds were not significantly less accurate than 6-year-olds ($t(81) = -0.97, p = 0.336$). These results indicate that response accuracy improved between 4 and 5 years of age and then remained at the same level at 6 years of age.

Importantly, the absence of an age x 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 Bonferroni-corrected, two-tailed, paired *t*-tests comparing accuracy on *aligned-same* and *misaligned-same* trials within each age group. As Figure 2 shows, accuracy on *aligned-same* trials ($M = 0.62$) was significantly lower than accuracy on *misaligned-same* trials in the 4-year-olds ($M = 0.72$; $t(48) = -3.03$, $p = 0.012$), 5-year-olds (*aligned-same* $M = 0.71$, *misaligned-same* $M = 0.88$; $t(46) = -6.93$, $p < 0.001$), and 6-year-olds (*aligned-same* $M = 0.74$, *misaligned-same* $M = 0.91$; $t(36) = -5.47$, $p < 0.001$).

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 (see Figure S2) yielded main effects of alignment ($F(1,127) = 7.95$, $p = 0.006$) and age ($F(2,127) = 20.22$, $p < 0.001$) but no effect of gender ($F(1,127) = 0.99$, $p = 0.321$) nor any interactions (age x alignment interaction: $F(2,127) = 0.79$, $p = 0.458$; gender x alignment: $F(1,127) = 0.00$, $p = 0.980$; age x gender: $F(2,127) = 0.328$, $p = 0.721$; age x gender x alignment: $F(2,127) = 0.85$, $p = 0.430$).

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 ($M = 0.58$) than 5-year-olds ($M = 0.76$; $t(94) = -3.41$, $p = 0.029$) and 6-year-olds ($M = 0.90$; $t(68) = -7.68$, $p < 0.001$) and that 5-year-olds were less accurate than 6-year-olds ($t(65) = -3.35$, $p = 0.003$). 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* than in the *misaligned-different* trials, these differences were not statistically significant at any age after correction (4-year-olds: *aligned-different* $M = 0.63$, *misaligned-different* $M = 0.54$; $t(48) = 1.83$, $p = 0.219$; 5-year-olds: *aligned-different* $M = 0.80$, *misaligned-different* $M = 0.72$; $t(46) = 2.27$, $p = 0.083$; 6-year-olds: *aligned-different* $M = 0.91$, *misaligned-different* $M = 0.89$; $t(36) = 0.63$, $p = 1.00$). 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 the *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* vs. *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 expectation that exposure to masked faces might have negative effects on the developmental emergence of holistic face processing. For this analysis, first we recoded all categorical variables as ordinal variables and dropped any “prefer not to answer” or non-responses, resulting in values for 11 COVID-19 questions from 96 participants. We then performed an exploratory factor analysis on the COVID-19 questionnaire data as a data-driven dimensionality reduction step (Fabrigar, Wegener, MacCallum, & Strahan, 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 1) was significantly different from the identity matrix ($\chi^2(950, 96) = 442.06, p < 0.001$), indicating that dimensionality reduction would be appropriate. Furthermore, the Kaiser-Meyer-Olkin test of sampling adequacy revealed a mediocre but acceptable value of 0.65. 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 2) revealed 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, while measures of the severity of masking in the community early in the pandemic loaded heavily onto the fourth latent factor.

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-squared revealed that this model only explained 3.34% of the variance in accuracy. Although age remained a significant predictor of accuracy ($b = 0.09, t(89) = 2.59, p = 0.011$; Table 3), none of the other factors contributed significantly to it (all $ps > 0.05$). 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 4. Neither age nor any of the latent factors from the COVID-19

questionnaire data predicted the difference in children's accuracy for *misaligned* minus *aligned-same* trial accuracy (all $ps > 0.10$). 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 to 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 a sub-sample of the children who were initially tested in Experiment 1 with identical but inverted faces. By testing the same children, we were able to control for between-subject variability.

Method

Participants. We re-contacted a subset of the participants ($N = 85$) from Experiment 1 to participate in a follow-up study that we conducted between December 2021 and February 2022. Out of the 85 contacted participants, 35 of them ($N = 15$ female gender; Mean age = 5.59, $SD = 0.88$; range: 4.19 to 7.15 years) 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. A binomial test revealed no difference in the gender distribution between Experiments 1 and 2 (47.3% non-male in Experiment 1 vs. 42.8% non-male in Experiment 2; binomial $p = 0.616$). Additionally, there was no difference in the children's average age between the two experiments (5.39 vs. 5.59 years; $t(164) = -1.19$, $p = 0.237$). The final sample consisted of 4-year-olds ($N = 12$; Mean age = 4.60 $SD = 0.21$; 2 females), 5-year-olds ($N = 9$; Mean age = 5.43, $SD = 0.24$; 5 females), and 6-year-olds ($N = 14$; Mean age = 6.55, $SD = 0.32$; 8 females). As in Experiment 1, the parents tended to identify as White (57.14%), Biracial (25.71%), or Asian (8.57%) from suburban (60.00%) or urban (37.14%) areas and with high levels of education (graduate/professional degree: 65.71%, Bachelor's degree: 28.57%) and wealth (families with annual income greater than or equal to 100,000 in: 65.71%).

During data acquisition for Experiment 2, a temporary error on the Lookit server caused spontaneous drop outs during the experiment. In the following analyses, we included all 35

participants who completed a full session of the experiment, regardless of whether or not they had previously attempted to participate. Because of randomization of the stimuli to different trial conditions for each participant, it is unlikely that the same test trials were administered to the participants in their final study attempt. Nonetheless, practice effects may have affected performance. Therefore, we explored this in Figure S3 by (a) restricting our analyses to participants who completed the experiment in one session ($N = 26$) 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 (Figure 1C).

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

Results and Discussion

Response Time

Same trials

The average response time was 3.01 seconds (range: 0.15 to 8.43 seconds). An ANOVA of response time in the *same* trials with alignment as a within-subjects factor, age and gender as between-subjects factors, and participant as a random effect yielded no main effect of alignment ($F(1,29) = 0.98, p = 0.331$), age ($F(2,29) = 0.16, p = 0.854$) nor gender ($F(1,29) = 0.03, p = 0.857$), and no interactions (age x alignment: $F(2,29) = 1.62, p = 0.216$; gender x alignment: $F(1,29) = 0.06, p = 0.806$; age x gender: $F(2,29) = 0.15, p = 0.862$; age x gender x alignment: $F(2,29) = 0.79, p = 0.462$).

Different trials

An ANOVA of response time in the *different* trials yielded no main effect of alignment ($F(1,29) = 0.06, p = 0.815$), age ($F(2,29) = 0.74, p = 0.487$), nor gender ($F(1,29) = 0.38, p = 0.543$), and no interactions (age x alignment: $F(2,29) = 1.16, p = 0.328$; gender x alignment: $F(1,29) = 0.33, p = 0.569$; age x gender: $F(2,29) = 0.40, p = 0.675$; age x gender x alignment: $F(2,29) = 0.09, p = 0.910$).

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 (since inversion is known to disrupt holistic processing). If, however, other factors such as response demands, executive control, 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 Figure 3 and, in contrast to Figure 2, accuracy was not lower in the *aligned-same* trials than in the *misaligned-same* trials. A mixed repeated-measures ANOVA on response accuracy yielded no main effects of alignment ($F(1, 29) = 0.40, p = 0.532$), age ($F(2, 29) = 2.06, p = 0.146$), gender ($F(1, 29) = 0.01, p = 0.912$), nor any interactions (age x alignment: $F(2,29) = 1.61, p = 0.217$; gender x alignment: $F(1,29) = 0.12, p = 0.733$; age x gender: $F(2,29) = 0.92, p = 0.411$; age x gender x alignment: $F(2,29) = 0.84, p = 0.443$). These results are consistent with our prediction that face inversion should interfere with the holistic face processing that the same children exhibited in Experiment 1.

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 ($F(2, 29) = 4.56, p = 0.019$) but no main effects of alignment ($F(1, 29) = 0.08, p = 0.775$) nor gender ($F(1, 29) = 0.05, p = 0.831$) and it did not yield any interactions (alignment x age: $F(2, 29) = 1.67, p = 0.205$; gender x alignment: $F(1, 29) = 0.63, p = 0.432$; age x gender: $F(2, 29) = 0.22, p = 0.803$; age x gender x alignment: $F(2, 29) = 1.65, p = 0.209$). Follow-up, two-sample Bonferroni-corrected, *t*-tests of response accuracy across age indicated that the 4-year-olds were less accurate ($M = 0.52$) than both 5-year-olds ($M = 0.81; t(18) = -2.99, p = 0.024$) and 6-year-olds ($M = 0.77; t(24) = -2.76, p = 0.033$) but that 5-year-olds did not differ from 6-year-olds ($t(18) = 0.42, p = 1.00$). Overall, this shows that the two older groups of children were more accurate than the youngest group of children in the inverted *different* trials.

Comparison of Experiments 1 and 2

Given that 35 of the children who participated in the current study completed both experiments, we had a sub-sample of participants who provided us with an important control for the possible contribution of individual differences to the data obtained in Experiment 1. As a result, we compared the performance of these 35 children across the two experiments. For this analysis, we collapsed the accuracy scores across the different age groups, primarily because we found no age x alignment interactions with *same* trial response accuracy in either Experiment 1 or Experiment 2 (Figure 4). We expected that accuracy would differ in the upright vs. inverted *aligned-same* trials but that it would not differ in the upright vs. inverted *misaligned-same* trials. Consistent with our prediction, paired *t*-tests (uncorrected) revealed that accuracy was lower in the upright ($M = 0.69$) than in the inverted ($M = 0.82$) *aligned-same* trials ($t(34) = -0.13, p = 0.009$), but that accuracy did not differ in the upright ($M = 0.84$) and inverted ($M = 0.85$)

misaligned-same trials ($t(34) = -0.01, p = 0.809$). Thus, overall, inversion had the predicted effect on accuracy in the *aligned-same* trials.

Discussion

The results from this experiment were consistent with our prediction that face inversion would disrupt holistic face processing. The same children who exhibited holistic face processing in Experiment 1 no longer exhibited it when the faces were inverted. This was evident in comparable accuracy scores in the *aligned-same* as in the *misaligned-same* trials.

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, using an online platform, we measured 4-, 5-, and 6-year-old children's ability to process faces holistically as well as their exposure to masked faces with a questionnaire administered 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, with lower accuracy scores in *aligned-same* than *misaligned-same* trials. Results also showed that accuracy scores were not correlated with degree of mask exposure. In Experiment 2, we replicated Experiment 1 except that this time we inverted the faces. As predicted, children no longer exhibited evidence of holistic face processing (i.e., lower accuracy scores in the *aligned-same* than in the *misaligned-same* test trials).

Our findings that 4-, 5-, and 6-year-old children exhibited holistic face processing replicate the 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 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 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 5-6 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 Ventura et al.’s may be that, as argued by some (Murphy, Gray, & Cook, 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 during early childhood (Gathercole, Pickering, Ambridge, & Wearing, 2004), the inclusion of an additional working memory component in the Ventura et al. study may have 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 makes it possible that young children’s holistic face processing is affected differentially by adult vs. children’s 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, 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 present 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 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 examining the statistical relationship between exposure and response accuracy in the composite face task. 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 deprivation of fully-visible faces to have some measurable effect. Therefore, we can only speculate on the reasons why we did not find a relationship between degree of exposure to masked faces and accuracy scores. It may be that holistic face perception is sufficiently robust by 4 years of age and, as a result, it is no longer vulnerable to disruption by the relatively short period of deprivation of fully-visible faces. Similarly, it may be that exposure to fully-visible faces at home was sufficient to overcome the deprivation outside the home. It may also be that exposure to partially visible faces in early childhood has relatively subtle effects and that more targeted measures, such as gaze/selective attention and/or neural markers of face processing, are needed to reveal the effects of the short-term deprivation. Neural markers may be an especially effective way to examine the effects of deprivation, given that some have found neural differences in infants' processing of face

identities pre- and post-lockdown (Yates, Ellis, & Turk-Browne, 2022). Finally, it may be that static face processing is relatively resistant to the effects of masking, but that dynamic face processing (particularly of talking faces) is more plastic with regards to experience. For instance, 10-12 month-old infants do not discriminate the faces of other races when tested with static faces, but do discriminate them when tested with dynamic faces (Minar & Lewkowicz, 2017).

One of the unique features of the current study is that it examined the development of the composite face effect outside 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). In general, there are several notable advantages to using an online platform. Data can be obtained from many children over a shorter period of time than in a laboratory-based study. Moreover, 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 socio-economically more diverse than the populations that participate in typical laboratory studies. The most likely reasons for this are that reaching more diverse populations may be limited by differential access that such populations have to the technology required to participate in an online study in general. Therefore, our results should be interpreted as reflecting a particular demographic. Indeed, it could be that face processing in certain groups of children may be more or less impacted 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 faces. Such faces might be appropriate for White children but not for children from other races or ethnicities. If holistic face processing depends, in part, on the specific early experience that children have with faces of a specific race or set of

racess, 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.

References

- 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.
- 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.
- Cassia, V. M., 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.
- Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379-383.
- 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. doi: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.
doi:10.3758/s13414-016-1194-4
- 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. doi: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), 272.
- 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.
- 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. doi: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.
- 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.

- Minar, N. J., & Lewkowicz, D. J. (2017). Overcoming the other-race effect in infancy with multisensory redundancy: 10–12-month-olds discriminate dynamic other-race faces producing speech. *Developmental Science*. doi:10.1111/desc.12604
- 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.
- Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, 31(5), 553-566.
- 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), e0185093. doi:10.1371/journal.pone.0185093
- Murphy, J., Gray, K. L., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review*, 24(2), 245-261.
- 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), 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., Haan, M. d., & 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.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing.

Psychological Bulletin, 140(5), 1281.

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. Paper presented at the Proceedings of the 9th Python in Science Conference.

Simion, F., & Di Giorgio, E. (2015). Face perception and processing in early infancy: Inborn

predispositions and developmental changes. *Frontiers in Psychology*, 6, 969.

Simion, F., Leo, I., Turati, C., Valenza, E., & Dalla Barba, B. (2007). How face specialization

emerges in the first months of life. In C. von Hofsten & K. Rosander (Eds.), *Progress in Brain Research* (Vol. 164): Elsevier.

- Simion, F., Valenza, E., Cassia, V. M., 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), 1-10.
- 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.
- Yates, T. S., Ellis, C. T., & Turk-Browne, N. B. (2022). Face processing in the infant brain after pandemic lockdown. *bioRxiv*.

Tables

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Child face perception	—										
2. Daycare status past	-0.10	—									
3. Daycare hours past	-0.06	0.79*	—								
4. Daycare status current	-0.13	0.36*	0.36*	—							
5. Daycare hours current	-0.09	0.46*	0.57*	0.76*	—						
6. Mandates past	-0.02	0.12	0.05	0.17	0.06	—					
7. Mandates current	-0.15	0.31*	0.19	0.47*	0.39*	0.37*	—				
8. Community mask past	-0.08	0.37*	0.31*	-0.04	0.10	0.36*	0.16	—			
9. Community mask current	-0.16	0.31*	0.30*	0.40*	0.38*	0.19	0.26*	0.48*	—		
10. Household mask past	-0.11	0.22*	0.19	0.08	0.08	0.35*	0.11	0.47*	0.19	—	
11. Household mask current	-0.08	0.25*	0.20	0.33*	0.29*	0.16	0.21*	0.29*	0.39*	0.62*	—

Table 1. Summary of Pearson correlations between COVID-19 variables in the questionnaire data for 96 of the 133 participants from Experiment 1. * $p < 0.05$, not corrected for multiple comparisons.

Measure	Factor			
	1	2	3	4
1. Child face perception	-0.14	-0.04	-0.06	-0.11
2. Daycare status past	0.25	0.73	0.10	0.19
3. Daycare hours past	0.21	0.96	0.08	0.03
4. Daycare status current	0.94	0.16	0.11	-0.02
5. Daycare hours current	0.71	0.42	0.10	-0.03
6. Mandates past	0.18	-0.05	0.08	0.59
7. Mandates current	0.52	0.07	0.01	0.34
8. Community mask past	-0.10	0.33	0.25	0.71
9. Community mask current	0.34	0.24	0.24	0.33
10. Household mask past	-0.02	0.11	0.59	0.41
11. Household mask current	0.25	0.08	0.96	0.10

Table 2. Factor loadings for the 11 different COVID-19 variables in a four-factor model with varimax rotation. The top two loadings for each factor are bolded to enable interpretation. Values represent standardized regression coefficients.

<i>Aligned-same trial accuracy by Factor 1 + Factor 2 + Factor 3 + Factor 4 + age + gender</i>				
	Beta estimate	Std. error	<i>t</i> -statistic	<i>p</i> -value
Factor 1 (Daycare current)	0.013	0.027	0.456	0.650
Factor 2 (Daycare past)	-0.050	0.028	-1.773	0.080
Factor 3 (Household masking)	-0.020	0.026	-0.773	0.441
Factor 4 (Community masking)	-0.014	0.031	-0.446	0.657
age	0.088	0.034	2.589	0.011
gender (non-male)	-0.009	0.054	-0.161	0.873

Table 3. Results from a general linear model predicting children’s accuracy on the aligned-same trials using the COVID-19 variable factors and age as predictors. Short descriptions of the four factors are provided to aid interpretation but are not exhaustive; see Table 2 for factor loadings. The equation used in the model is shown as the table heading. The beta parameter, standard error, *t*-statistic, and *p*-values are given for each of the predictors.

(*Misaligned-same* - *aligned-same* trial accuracy) by Factor 1 + Factor 2 + Factor 3 + Factor 4 + age + gender

	Beta estimate	Std. error	<i>t</i> -statistic	<i>p</i> -value
Factor 1 (Daycare current)	-0.018	0.024	-0.729	0.468
Factor 2 (Daycare past)	0.025	0.025	1.002	0.319
Factor 3 (Household masking)	-0.003	0.023	-0.147	0.883
Factor 4 (Community masking)	0.043	0.027	1.617	0.109
age	0.022	0.030	0.738	0.462
gender (non-male)	0.023	0.047	0.493	0.623

Table 4. Results from a general linear model predicting the difference in children's accuracy on the *aligned-same* vs. *misaligned-same* trials using the COVID-19 variable factors and age as predictors. Short descriptions of the four factors are provided to aid interpretation but are not exhaustive; see Table 2 for factor loadings. The equation used in the model is shown as the table heading. The beta parameter, standard error, *t*-statistic, and *p*-values are given for each of the predictors.

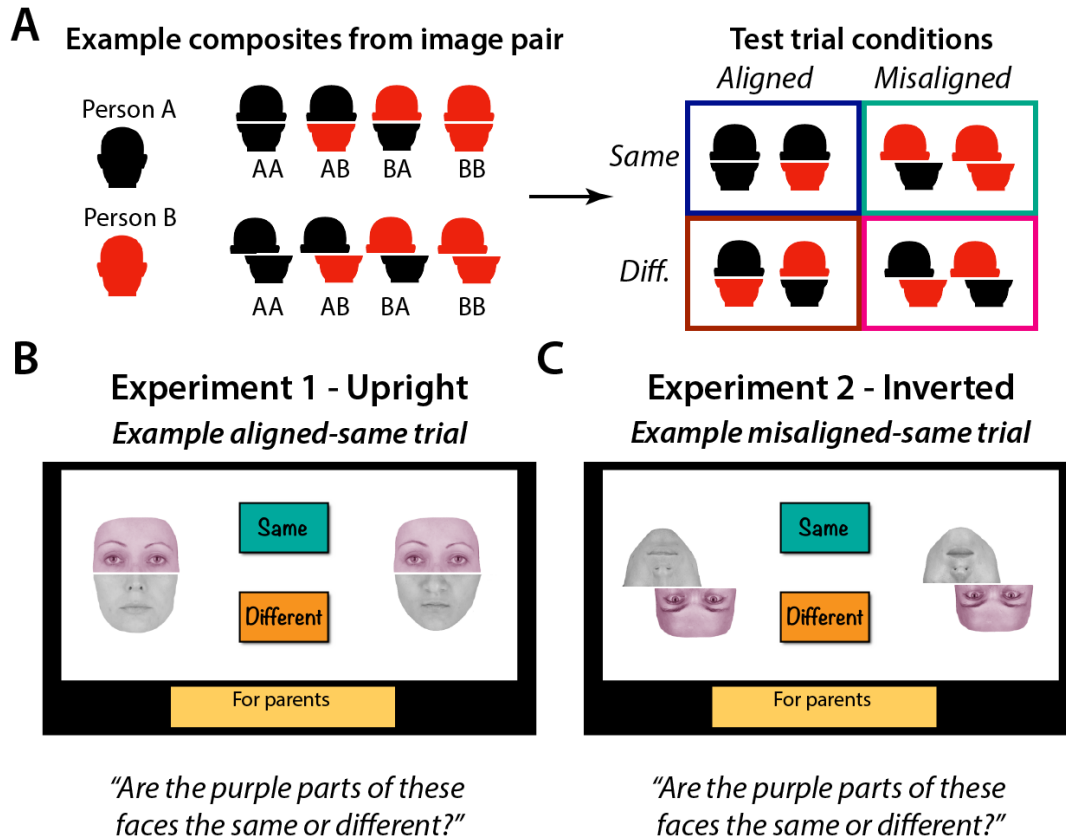


Figure 1. Stimuli presented in Experiments 1 and 2. (A) Cartoonized example of how face identity pairs were combined to make different composite faces. On a given test trial, participants saw composite faces presented on the left and right sides of the screen according to 4 different conditions: aligned-same,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

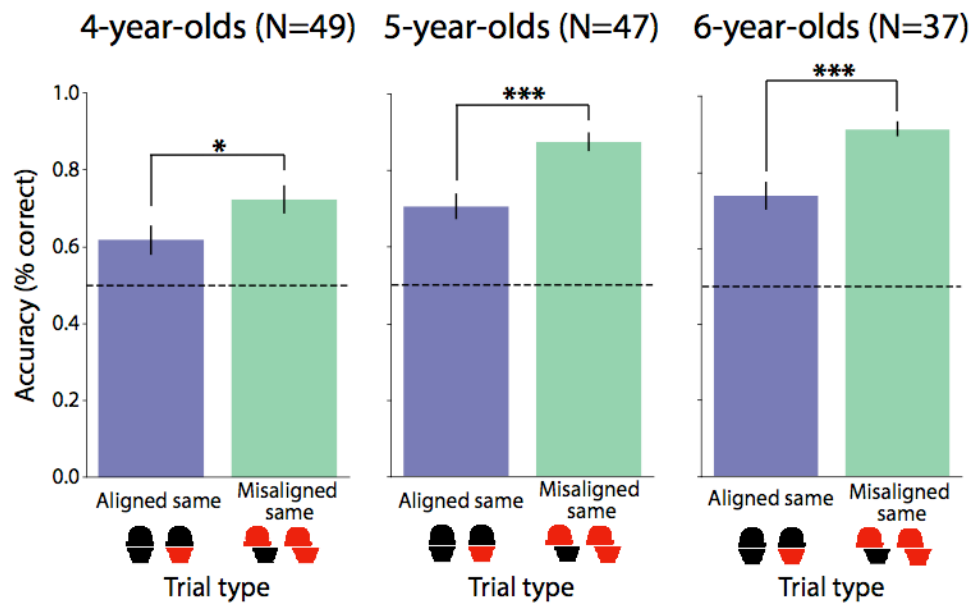


Figure 2. Accuracy in the *aligned-same* and *misaligned-same* trials as a function of age in Experiment 1. Asterisks denote Bonferroni-corrected significant differences (*** $p < 0.001$, * $p < 0.05$) and error bars represent standard errors of the mean.

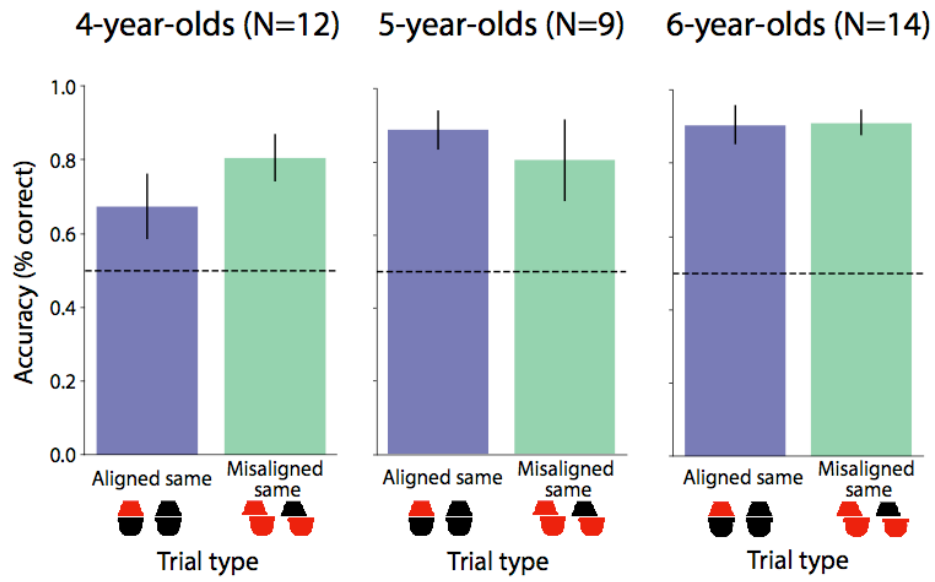


Figure 3. Accuracy on the *aligned-same* and *misaligned-same* trials as a function of age in Experiment 2. Error bars represent standard error of the mean.

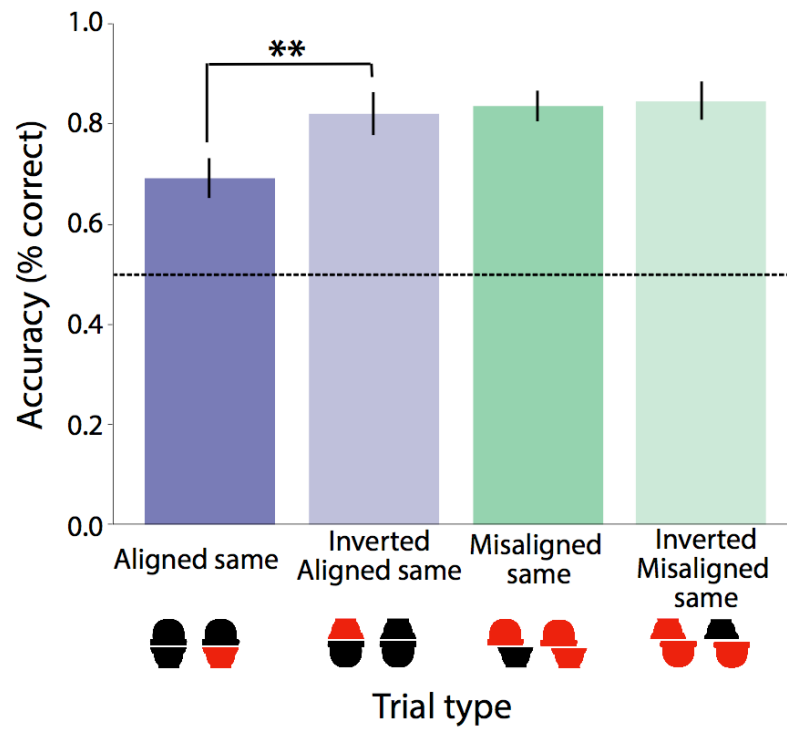


Figure 4. Accuracy in the *aligned-same* and *misaligned-same* trials for upright vs. inverted composite faces in Experiments 1 and 2. Error bars represent standard errors of the mean. ** $p < 0.01$.