Configural properties of face portraits change between childhood and adulthood

Balas, Benjamin; Weigelt, Sarah; Koldewyn, Kami

International Journal of Behavioral Development

Accepted/In press: 24/05/2022

Peer reviewed version

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Balas, B., Weigelt, S., & Koldewyn, K. (Accepted/In press). Configural properties of face portraits change between childhood and adulthood. *International Journal of Behavioral Development*.

Adult observers are sensitive to the configuration of facial features within a face, able to distinguish between relative differences in feature spacing, and detecting deviations from typical facial appearance. How does the representation of the typical configuration of facial features develop? While there is a great deal of work describing children's developing abilities to detect differences in feature spacing across face images, there is substantially less work examining what children think constitutes a typical arrangement of facial features. In the current study, we investigated this issue using a production task in which adults and 5-10 year-old children created a face "portrait" by arranging the eyes, nose, and mouth of a standard face within an empty outline. Using this simple task, we found differences in face configuration across age groups, such that children of all ages made far larger errors than adult participants, expanding facial features outward from the center of the face more than adults. These results were not affected by face inversion, potentially implying a domain-general rather than face-specific process. We also found that children of all ages endorsed the correct configuration as a best likeness in a perceptual task. We discuss these results in terms of ongoing debate regarding the extent to which configural processing is a meaningful component of face recognition, and the conclusions we can draw from production paradigms as compared to purely perceptual tasks.

Keywords: Face recognition; visual development; configural processing; drawing

Introduction

Though adults are typically very sensitive to distortions of facial appearance that affect the geometry or configuration of features in perceptual tasks, they also make large, systematic errors in configuration when drawing or otherwise producing a face image. In particular, adults tend to suffer from a robust error in the placement of the eyes within the facial outline such that they are placed much too high - a phenomenon that has been called "The Squashed Skull Effect" (Edwards, 1999). This bias is evident when adults create face images from memory or by copying a face image (Carbon and Wirth, 2014), but can be reduced by training in portraiture (Ostrofsky et al., 2016). This discrepancy between the production of face images (prone to errors in eye placement that are large in magnitude per Ostrofsky (2013)) and the perception of face configuration in adulthood is intriguing and suggests a potentially important dissociation between mechanisms for recognizing faces and mechanisms for creating face images with appropriate spatial relationships. Examining face drawings, or other face images created by naïve observers, is thus interesting to examine in its own right: The discrepancies between errors that are tolerated when making an image vs. when viewing an image demonstrate that drawing is governed by unique processes that are worth investigating on their own both in adults and developmentally. Further, though production abilities do not seem to be entirely constrained by perceptual processing, drawing provides an opportunity for observers to use their own perception of the image they are creating to guide the process of creation. Developmentally, this means that while discrepancies between production and comprehension or perception exist in other domains, drawing is potentially unique in that drawings can be erased and edited to more closely align with an individual's perception of what is appropriate. Our goal in the current study was to examine face production and perception in middle childhood (5-10 years of age) relative to adulthood in order to examine how production/perception discrepancies may change in magnitude with age and also to examine face production errors developmentally more broadly than previous studies have done.

Adult-like perception of facial configuration develops relatively slowly during childhood. Though in some instances children younger than 6 or 7 years of age behave in adult-like ways with regard to configural information in face images (McKone and Boyer, 2006; Pellicano, Rhodes, and Peterson, 2006), children's ability to measure spatial relationships between faces matures during middle childhood. Mondloch, Leis, and Maurer (2006), for example, found that 4-year-olds were generally poor at detecting spacing changes made to familiar faces they were exposed to in storybook, despite evidence that preschool-age children can use face configuration for identification in some limited cases (Freire and Lee, 2001). In a different study considering a wider age range, Mondloch et al. (2002) found that children across the 6-10 year age range were worse than adults at detecting differences in feature spacing, but matched adults performance for changes in the external contour of the face and the shape of local features at different points in this age range. This result demonstrates that children are not simply less efficient at processing facial appearance considered broadly, but instead have a distinct developmental trajectory for configural processing of facial features. An important alternate view, however, is that a number of these results may reflect methodological issues related to the matching of task difficulty across featural and configural changes to the face and that removing these confounds reduces the extent to which development change is observed in this age range that is specific to configural aspects of face appearance (McKone et al., 2012) Configural processing and its development are not limited to the evaluation of spacing relationships in facial patterns either. The ability to detect the grotesqueness associated with Thatcherization of face patterns is not yet adult-like in childhood (Donnely and Hadwin, 2003; Rouse et al., 2004), though the dependence of the phenomenon on inversion appears to be stable from the age of 6 onwards (Lewis, 2003). Mondloch et al. (2004) also found that 8-year-old children reported reduced "bizarreness" ratings relative to adults when asked to evaluate faces that were severely distorted by changes to feature spacing or Thatcherization, but approximately matched adult ratings when evaluating changes in local feature appearance. Children therefore appear to need extended development during middle childhood to achieve adult-like sensitivity to differences in feature spacing, and are also not as sensitive as adults to violations of typical face configuration.

Given that adult-like perception of face configuration emerges slowly during childhood in some tasks, how do children's face production abilities change during the same period? With regard to the eye placement errors specifically, it appears that older children make smaller errors than younger children, ultimately converging on the same systematic error that adults make. Using a unique stimulus set of self-portraits made on tea-towels for a classroom project, McManus et al. (2012) determined that children's errors in eye placement lessened between the ages of 3 and 11. Using a more constrained task with schematic faces, Smith, Kempe and Wood (2021) similarly demonstrated improvements in the eye placement bias as age increased via a comparison of 3-5 year-old children, 10-11 year-old children and adults and also reported larger errors when faces were created from memory. Their paradigm is particularly useful in that it minimizes the possible contribution of motor development to these errors by asking children to arrange parts within a face outline rather than draw a face with a stylus or marker. Children's production and perception of face images thus both continue to develop during middle childhood, with face production ultimately incorporating at least one source of systematic error in the form of the eye placement bias described above.

Presently, we examined children's face production abilities to build on prior research in several distinct ways. First, as we have described above, prior reports regarding children's production of face images have largely focused on the vertical placement of the eyes within the facial outline, which is only one aspect of face geometry. One of our goals in the current study was thus to determine if there are consistent placement biases that affect the position or spacing of other facial features, or if the eyes, nose, and mouth may be more or less subject to error during development. Second, we also investigated the role of face inversion on face production in the current study to determine how a manipulation that profoundly affects the perception of face configuration may affect the creation of face images in children. While adult observers create portraits that are less accurate when subject to face inversion (Day and Davidenko, 2018), to our knowledge the effect of inversion on children's drawings have not been studied. Finally, we present our production task to children in the context of a typicality judgment rather than as a memory or copying task involving a specific face. Drawing (and other kinds of production task) are useful tools for examining internal representations of objects at multiple levels of specificity (Kosslyn, Heldemyer and Locklear, 1977; Long, Fan and Frank, 2018). Rather than probing the nature of a child's prototype through manipulations of multiple stimuli, production tasks make it possible to ask the participant to show you their estimate of an object's appearance directly. As such, our study is an attempt to not only comment on factors affecting errors in feature placement in children's face portraits, but also to reveal properties of their emerging representation of typical face appearance. Children's understanding of typical appearance is also changing during middle childhood, as evidenced by tasks measuring children's preferences for typical faces over distorted counterparts (Cooper et al, 2006; Short et al., 2015). The results of these studies (which we describe in more detail in our discussion section below) combined with our production data reveal a complex relationship between face perception and production that intersects with emerging sensitivity to specific metric relationships in face images and face norms that are continuing to be refined based on experience.

In our first task, we asked adults and children between the ages of 5 and 10 years of age to create face "portraits" by positioning the eyes, nose, and mouth of a standard face within an outline. Compared to a fully unconstrained drawing task, this production task limits the participant a great deal: The only degrees of freedom they can meaningfully explore are the 2D positions of the features within the outline. Feature shape and size, line width, and other aspects of a drawing are fixed by our choice of testing materials. While these constraints limit the range of behavior we can expect from our participants by limiting the richness of their creations, that also helps limit the analytical choices available to us as we characterize participants' responses. Allowing participants to create unconstrained drawings generally requires the development of a coding scheme to objectively describe each drawing, as well as reliability checks to ensure that coders appropriately described each participant response (#REF). By constraining our participants this way, we also obviated the need to develop such a scheme. All observers were asked to make a portrait that they thought reflected what a normal, or typical, face looked like. We asked different groups of participants to complete this task with upright and inverted materials so that we could examine how observers' estimate of typical face configuration may be affected by the orientation of the face (Baudouin et al., 2010). Following this, we conducted a perceptual task (Exp. 2) designed to reveal the perceptual abilities of participants in the target age range relative to the production abilities of their peers. Adopting the same framework as Balas and Sinha (2008) we asked a new group of children to choose the most typical face out of a lineup including the true population norm and the average portraits created by children in each age group. Overall, we expected to reproduce the previously reported effects of age on eye placement errors, such that younger children would make more dramatic errors in vertical placement than older children and adults. We also predicted that other systematic spacing differences would differentially affect the nose and mouth spacing as a function of age, with all of these errors being subject to an inversion effect that would further increase errors. Finally, we predicted that if children's perception and production abilities are more closely yoked than those of adults, children may be more likely to perceive the average face created by their age group as "typical" in our perceptual task, selecting that average over the true population norm. Alternatively, if production and perception are not tied to one another during childhood, we expected children may perform accurately in selecting the population norm, but perhaps only at later ages.

Experiment 1 – Upright and Inverted Face Production

Method

Participants

We recruited a total of 175 participants to create face portraits for this study. Approximately half of these participants were asked to make upright portraits (N=90) and the remaining participants were asked to make inverted portraits of a face (N=85). Within each orientation condition, we tested participants in 4 non-overlapping age groups: 5-6 year-olds ($N_{upright}$=24 (12M, 12F) , $N_{inverted}$=20 (7M, 7F)); 7-8 year-olds ($N_{upright}$=22 (5M, 17F), $N_{inverted}$=22 (9M, 13F)); 9-10 year-olds ($N_{upright}$=20 (8M, 12F), $N_{inverted}$=23 (12M, 11F)); and adults ($N_{upright}$=24 (12M,12F), $N_{inverted}$=20 (9M, 11F)). The average of the adults was 28 years old (S.D. = 4 years). Each participant group was predominantly White, with no more than 5 individuals self-identifying as non-white in any group. The experiment was also conducted in

North American cities with majority White populations. As this does not offer us adequate power to comment on the effects of participant race on performance we do not examine this variable further. With regard to overall statistical power for our factors of interest, a post-hoc power analysis carried out in G*Power 3.1 indicates that this sample size provides us with ~90% power (1-β) to detect an effect size of 0.25. All recruitment and testing procedures were approved by the NDSU IRB (#SM19258) in accordance with the principles outlined in the Declaration of Helsinki.

Stimuli

We created a “construction kit” for creating a face portrait using the Farkas norms (Farkas, 1994; Farkas, Hreczko and Katic, 1994) for the average facial geometry of a White, North American adult man. Though the choice of facial norm could affect subsequent error coding if participants have biased experience with faces of different races or different amounts of exposure to male vs. female faces, we chose to use the Farkas norm as it remains one of the most comprehensive databases of craniofacial anthropometry (Deutsch et al., 2012). We will revisit the potential issue of varying experience with different face categories in the Discussion section, but for now simply highlight this point for consideration. We created an external outline and appropriately sized eyes, nose, and mouth features (Figure 1) that were printed actual size on an 8”x11.5” piece of paper. The external contour of the face was printed out with the internal features removed to serve as a template for the placement of the internal parts. The internal features of the face were printed out on smaller pieces of paper so that they could be easily manipulated and taped down.

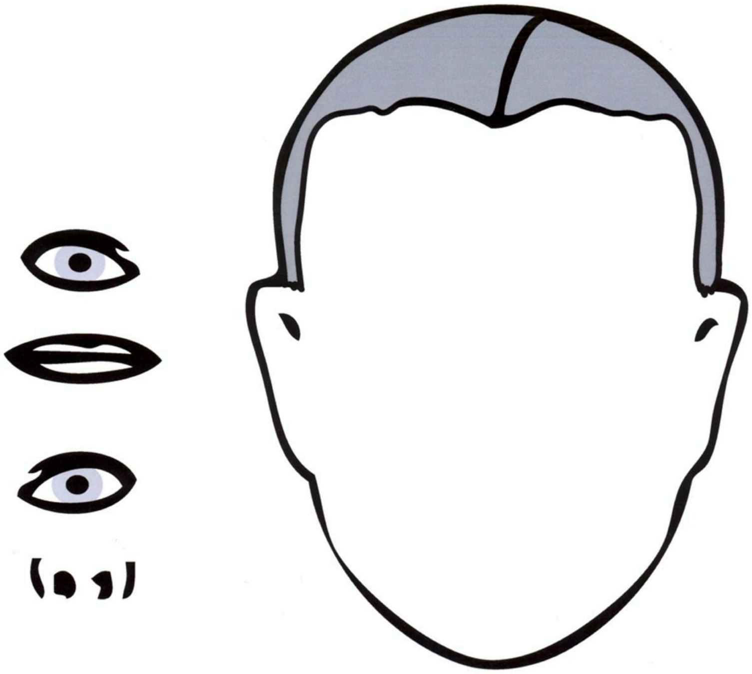

Figure 1. A schematic view of the production task used in our study. Children and adults were asked to place the features at left inside the empty face outline.

Procedure

Participants were given a piece of paper depicting the external contour of the face and were asked to place the internal features inside this template so that it looked like a typical face. Participants were told that this face was meant to be a man's face and were given as much time as they wished to complete the task. The external contour and the internal features were placed on a table under natural lighting during task completion, and the internal features were taped to the template once the participants indicated that they were satisfied with their portrait. Participants in the upright condition were presented with the external contour in a normal, upright orientation, while their counterparts in the inverted condition were presented with the contour upside-down on the table. Participants were not permitted to manipulate the orientation of the external contour, nor were they allowed to move around the table to change their viewing angle. Further, participants were discouraged from observation of a real face during completion of the task and did not have access to a mirror, camera, or other device for providing themselves with a view of their own face. Participants typically completed the task in less than 5 minutes.

Results

We chose to use two sets of descriptors to quantify the arrangement of facial features within each portrait made by our participants: Local inter-feature distances between discrete face parts (e.g. inter-ocular distance) and global estimates of the difference between each portrait and a standard template based on population norms (Procrustes distance). Pairwise feature spacing has been manipulated in many studies of face processing with children and adults, which motivated our choice to use these local descriptors to characterize our participants' portraits. The Procrustes distance, which we describe below in more detail, has to our knowledge not been applied to studies focused on characterizing sensitivity to configural information in face image. However, we think it offers a highly useful way to characterize the global layout of features within the face pattern.

We calculated both sets of descriptors using the x-y coordinates obtained for each of the four facial features participants were free to manipulate during the task. This coding did not include information about the orientation of the individual features (e.g. planar rotation of the eye), so our analyses do not speak to these aspects of facial appearance. We also note that qualitative errors (e.g., placing an eye in the position for the nose) were rare: Only two participants made such errors in feature placement and their portraits were excluded on this basis. We extracted the necessary measurements by identifying fiducial points for the two eyes, the nose, and the mouth on each portrait using Adobe Illustrator, and recording the x-y coordinates of each point. To ensure good registration between portraits, each raw portrait image was scaled to a standard size by adjusting the portrait size to match digital landmarks created using our original face template. This procedure ensured that there were not distortions in our measurements due to variation introduced during image scanning and importing. Summary figures depicting the placement of the eyes, nose, and mouth for each age group and orientation condition are displayed in Figure 2.

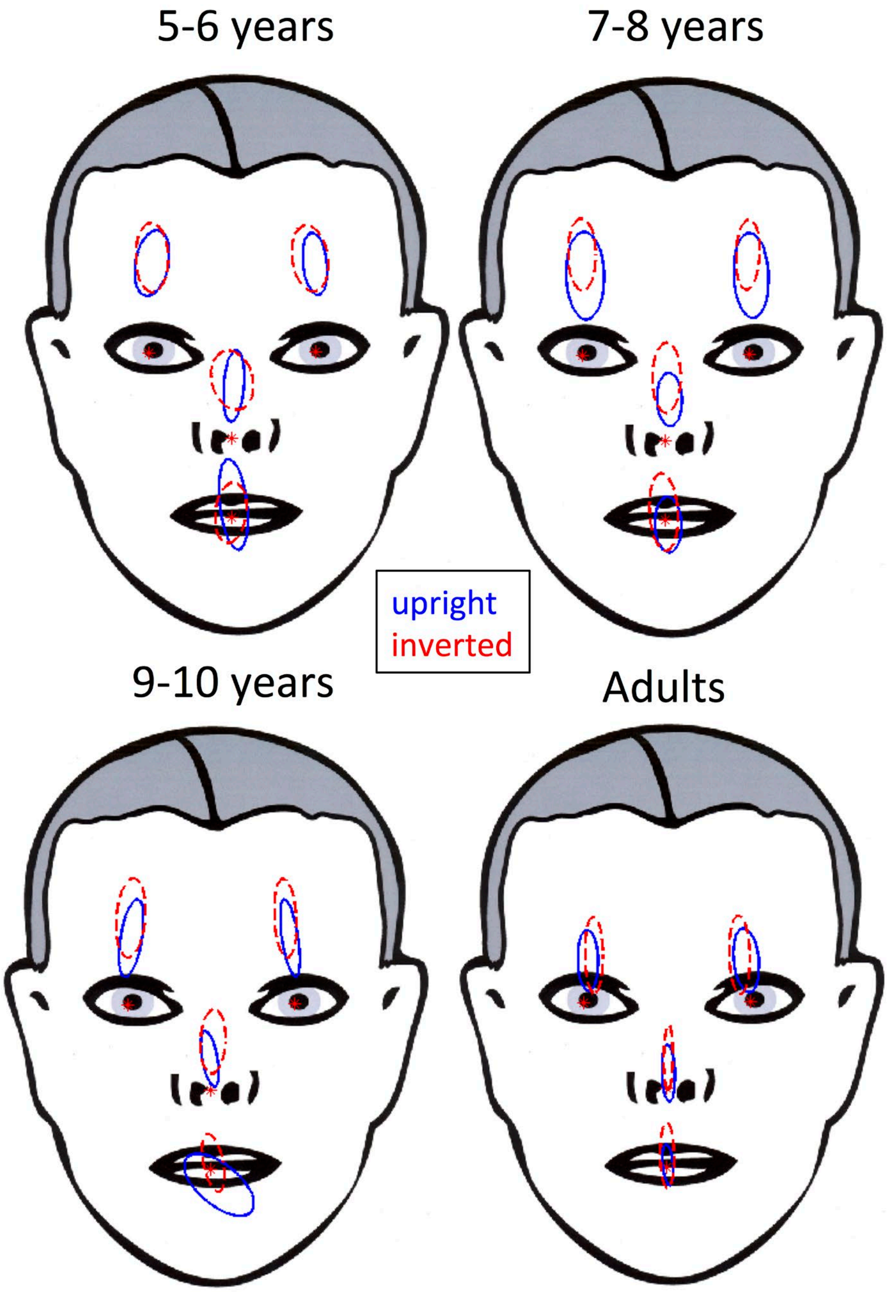

Figure 2. Average feature placement as a function of age and image orientation. Each ellipse (solid for upright, dashed for inverted) represents a 95% confidence interval for the x-y position of each feature within the standard outline. The asterisks depict the approximate best position for each feature. Sample sizes for

each age group are as follows: 5-6 y.o., N=24; 7-8 y.o., N=22; 9-10 y.o., N=20; Adults, N=20).

Global Descriptors (Procrustes Analysis)

Our global descriptors were obtained by calculating the Procrustes distance between the four points obtained from the portrait and the position of those same features in the original template. Briefly, the Procrustes distance is a means of calculating the distance between two sets of points following a linear transformation that is meant to bring the two sets of points into the closest possible alignment before the distances between corresponding points are calculated. Specifically, one set of points may be translated, rotated, and scaled to match the second set of points as closely as possible, after which the "stress" or residual error between the positions of corresponding points is calculated (Kendall, 1989). We selected this distance metric because it provides a way to consider the global arrangement of a set of multiple points at once. If we were to consider the position of each individual feature singly, for example, we would have no way to account for the joint statistics of position across the eyes, nose, and mouth. Because the Procrustes distance considers the entire set of points as one entity, we obtain a measure of distance that does respect that joint location information.

We measured the Procrustes distance using the function Procrustes.m in MATLAB, which allows the user to extract the residual distance between the two sets of points (which we will refer to here as '$D_{Procrustes}$') as well as the translation, scaling, rotation, and reflection coefficients used to bring the two sets of points into alignment. We chose not to examine the translation, rotation, or reflection components of the transform, primarily because we did not expect much variation in these parameters across participants. All participants tended to respect the vertical symmetry of the face pattern, for example, making both the rotation and reflection components near zero in all cases. Similarly, the translation vector required to bring the two sets of points into alignment was generally quite small. We therefore only examined the scaling coefficient (which we will refer to here as $B_{Procrustes}$) and the residual distance to characterize how face portraits changed developmentally. Briefly, the scaling coefficient describes how much one set of points had to expanded or contracted to best match another: A number greater than 1 indicates that expansion was necessary to bring the two sets of points into the closest possible alignment, while a number less than 1 indicates that compression was required. This value thus indicates whether our participants tended to bunch features up too close together relative to our norm, or whether they tended to spread them apart too much. The residual distance describes the remaining error beween the two sets of points after scaling, translation and rotation bring them into the closest possible alignment. Higher values indicate larger differences between the norm and the participant-generated points. Such residual errors are easiest to think about in this task as errors of proportion. For example, if the ratio between the eye-nose vertical distance and the nose-mouth distance is supposed to be approximately 1:1, but a participant instead creates a portrait with a ratio of 2:1, this difference in proportion cannot ever be corrected by uniform scaling, translation, or rotation. Similarly, if a single eye is misplaced horizontally while the other features are in the correct place, these points cannot all be brought into alignment via the transformations permitted when computing the Procrustes distance.

For each descriptor, we analyzed the results using a 4x2 Bayesian ANOVA carried out in JASP (JASP, 2018). Both participant age group (5-6 years old, 7-8 years old, 9-10 years old, and adults) and portrait orientation (upright, inverted) were between-participants factors. We continue by describing the outcome of each of these analyses in detail.

Global Scale ($B_{Procrustes}$). This analysis revealed very strong evidence in support of a main effect of Age Group ($BF_{10}=3.8 * 10^7$) such that scale coefficients increased with age (Figure 3). Specifically, adults' scale coefficient values were larger than those of each child group, though all groups exhibited values of $B_{Procrustes}$ that were smaller than 1. This reflects the fact that all participant groups spread facial features out more than necessary, taking up more of the face contour than is typical. Regarding orientation, we also observed strong evidence in support of the null hypothesis of no effect of inversion ($BF_{10}=0.18$), suggesting that inversion did little to change this feature of face portraits constructed by our participants. Finally, we examined the potential for an interaction between our factors by considering the ratio of Bayes Factors described in our previous analyses, revealing moderate support for the model that did not include an interaction term ($BF_{10}=0.4$).

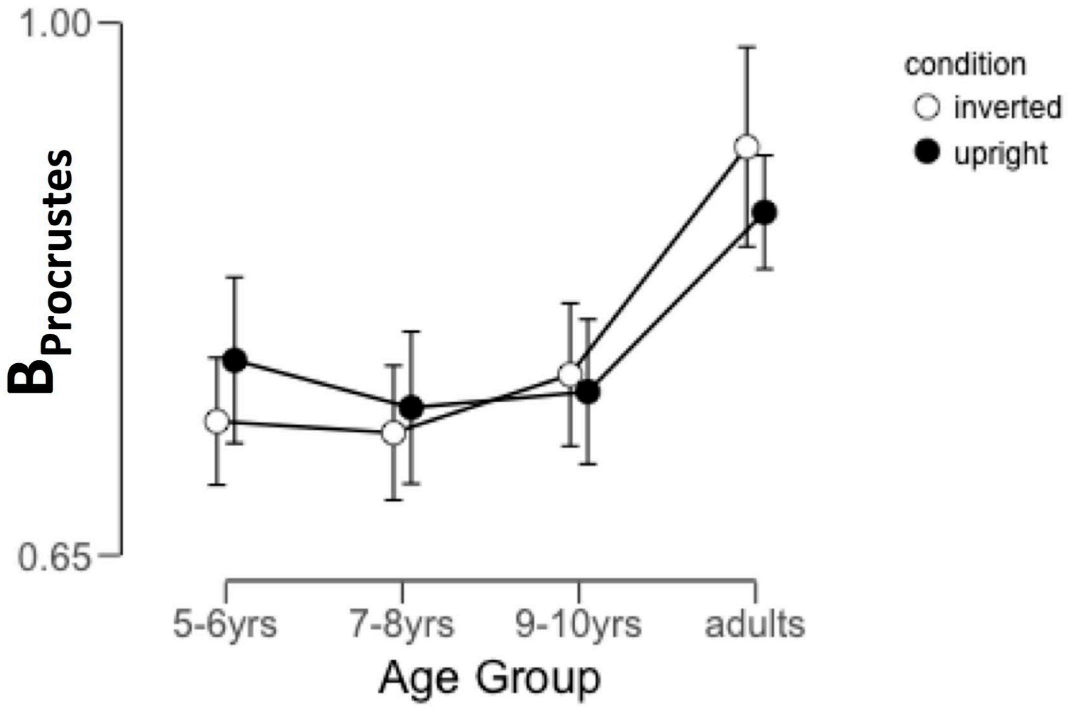

Figure 3. Average values of the scale coefficient from the Procrustes transformation aligning participant portraits with a standard template as a function of Age Group and portrait orientation. Error bars indicate 95% credible intervals. Sample sizes for each age group are as follows: 5-6 y.o., N=24; 7-8 y.o., N=22; 9-10 y.o., N=20; Adults, N=20).

Residual Error ($D_{Procrustes}$). Finally, we examined the residual global error in feature position after allowing for translation, scaling, rotation, and reflection of the points corresponding to participants' placement of the eyes, nose, and mouth. This analysis revealed strong evidence in support of a main effect of Age Group (BF_{10}=68.8), such that adults tended to have lower values of $D_{Procrustes}$ than children

(Figure 4). We found weak evidence in support of a null effect of orientation on the error term (BF_{10}=0.75), once again suggesting that orientation has little to no impact on this property of facial feature placement. Finally, we examined the potential interaction between these two factors according to the procedures described above, revealing a Bayes Factor of approximately 0.8, which offers weak support for the model that does not include an interaction term.

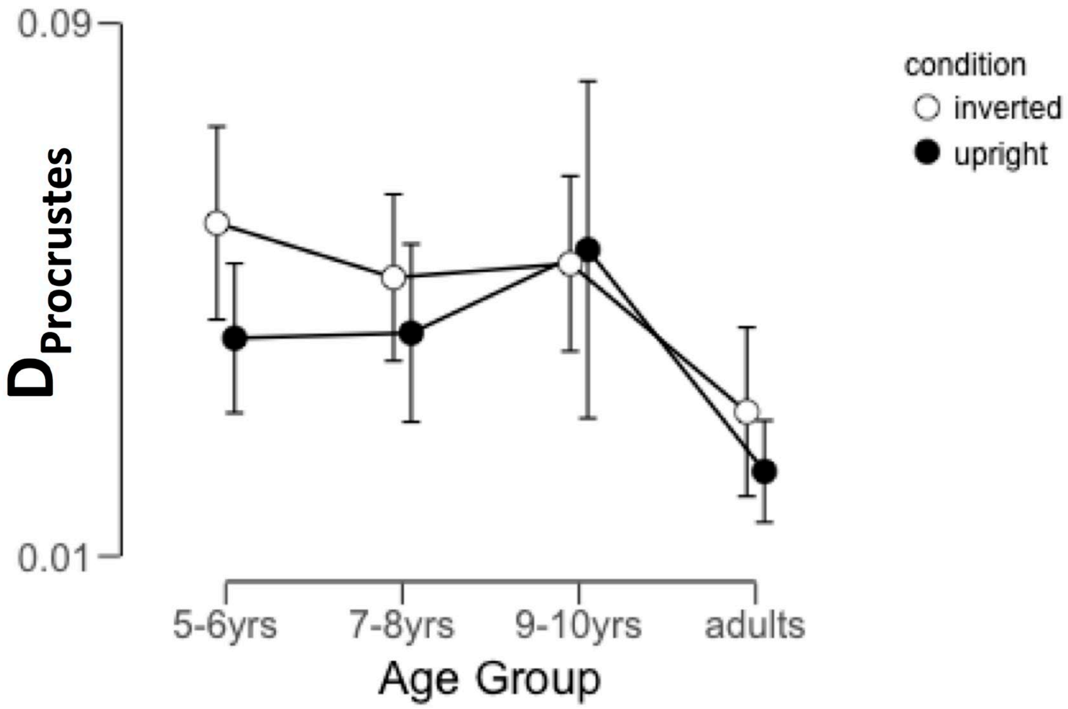

Figure 4. Average residual error ($D_{Procrustes}$) as a function of Age Group and portrait orientation. Error bars indicate 95% credible intervals. Sample sizes for each age group are as follows: 5-6 y.o., N=24; 7-8 y.o., N=22; 9-10 y.o., N=20; Adults, N=20).

Local inter-feature distances

Inter-feature distances (interocular distance, eye-nose distance, and nose-mouth distance) were each calculated by taking the difference between the relevant coordinates of the target features for each descriptor. For example, the interocular distance was calculated using the difference between the x-coordinates of the left and right eyes, while the vertical nose-mouth distance was calculated using only the y-coordinates of the nose and mouth.

As above, we analyzed the impact of age and orientation on portrait appearance using a 4x2 Bayesian ANOVA implemented in JASP, with age group and portrait orientation as between-subject factors.

Interocular Distance. Our analysis revealed substantial evidence in support of a main effect of age group on interocular distance (BF_{10}=45.3), such that the interocular distance tended to decrease with age (Figure 5). Post-hoc tests revealed that 7-8 year-olds differed from adults (BF_{10}=15.5), while 9-10 year-olds did not (BF_{10}=0.40). There was also only weak evidence supporting a difference between 5-6 year-olds and adults (BF_{10}=1.74), which may reflect the larger variability in this child age group. There was only weak evidence supporting a main effect of orientation, however (BF_{10}=1.42), such that inverted portraits tended to have smaller interocular distance values. We note that in general, participants tended to overestimate the interocular distance of typical faces, placing the eyes slightly further apart than normal. Finally, to examine the potential interaction between Age Group and Orientation, we calculated the ratio of the Bayes Factor associated with a model incorporating both main effects and an interaction term to the Bayes Factor associated with a model incorporating only both main effects. This yielded a Bayes Factor of approximately 0.15, which indicates strong support for the model that did not include an interaction term.

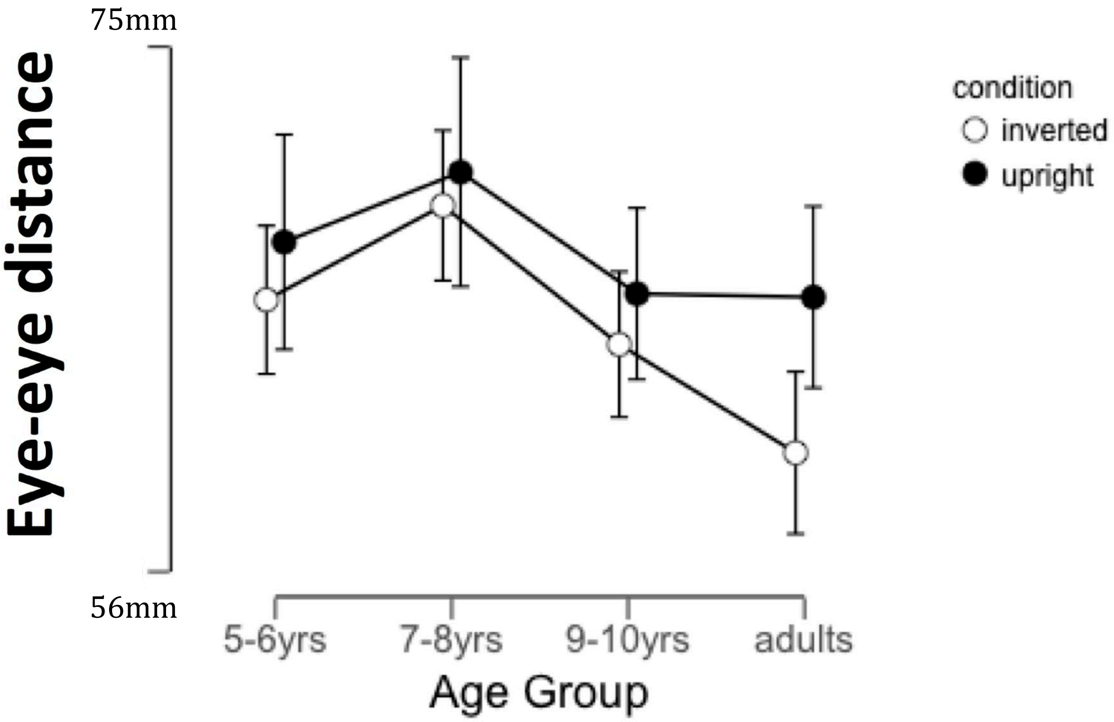

Figure 5. Average inter-ocular distance in pixels as a function of Age Group and portrait orientation. Error bars indicate 95% credible intervals. Sample sizes for each age group are as follows: 5-6 y.o., N=24; 7-8 y.o., N=22; 9-10 y.o., N=20; Adults, N=20).

Vertical Eye-Nose Distance. This analysis revealed moderate evidence in support of a main effect of Age Group on the vertical distance between the eyes and the nose

(BF_{10}=2.31), such that adults tended to make portraits with a smaller distance value than children of all ages (Figure 6). With regard to orientation, we observed strong evidence supporting the null hypothesis of no effect of inversion (BF_{10}=0.18), suggesting that orientation did not affect the relative vertical positions of these two features. Unlike interocular distance, participants also did not tend to overestimate this value in their portraits. Finally, to examine the potential interaction between Age Group and Orientation, we calculated the ratio of the Bayes Factor associated with a model incorporating both main effects and an interaction term to the Bayes Factor associated with a model incorporating only both main effects. This yielded a Bayes Factor of approximately 0.10, which indicates strong support for the model that does not include an interaction term..

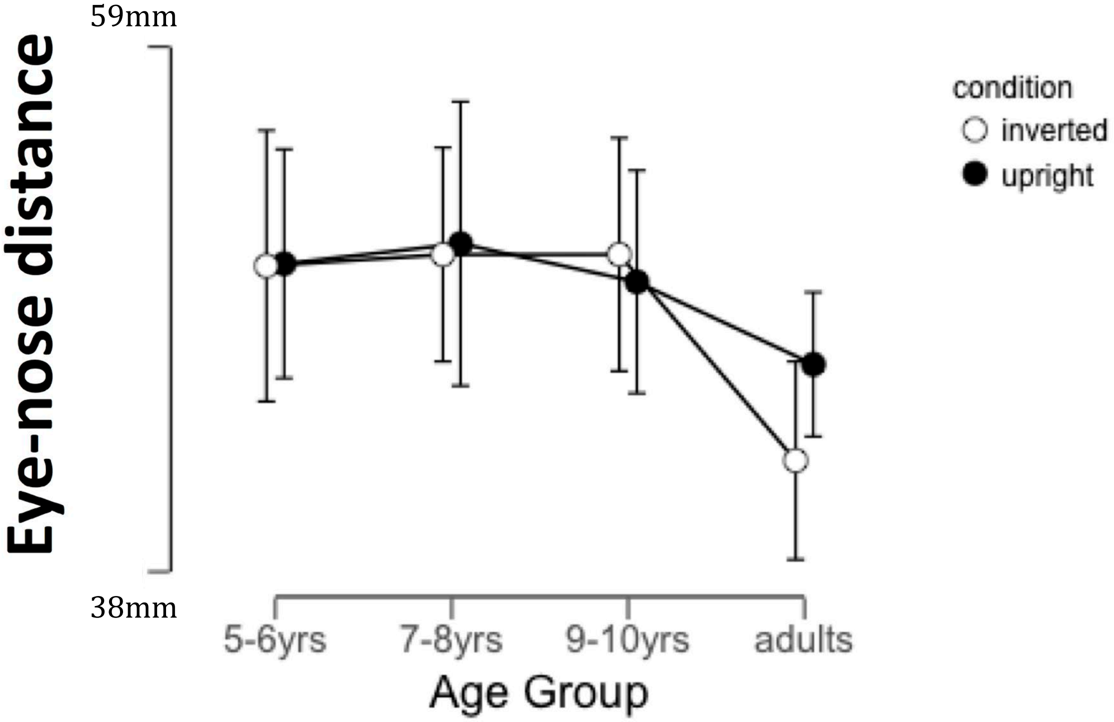

Figure 6. Average vertical distance in pixels between the eyes and nose as a function of Age Group and portrait orientation. Error bars indicate 95% credible intervals.

Vertical Nose-Mouth Distance. This analysis revealed very strong evidence in support of a main effect of Age Group on the vertical distance between the nose and the mouth ($BF_{10}=1.08 * 10^7$), such that adults tended to make portraits with a smaller distance value than children of all ages (Figure 7). With regard to orientation, we observed strong evidence supporting the null hypothesis of no effect of inversion, ($BF_{10}=0.55$), suggesting that orientation also did not affect the relative vertical positions of these two features. Finally, to examine the potential interaction between Age Group and Orientation, we calculated the ratio of the Bayes Factor associated with a model incorporating both main effects and an interaction term to the Bayes Factor associated with a model incorporating only both main effects. This yielded a Bayes Factor of approximately 0.2, which indicates strong support for the model that does not include an interaction term.

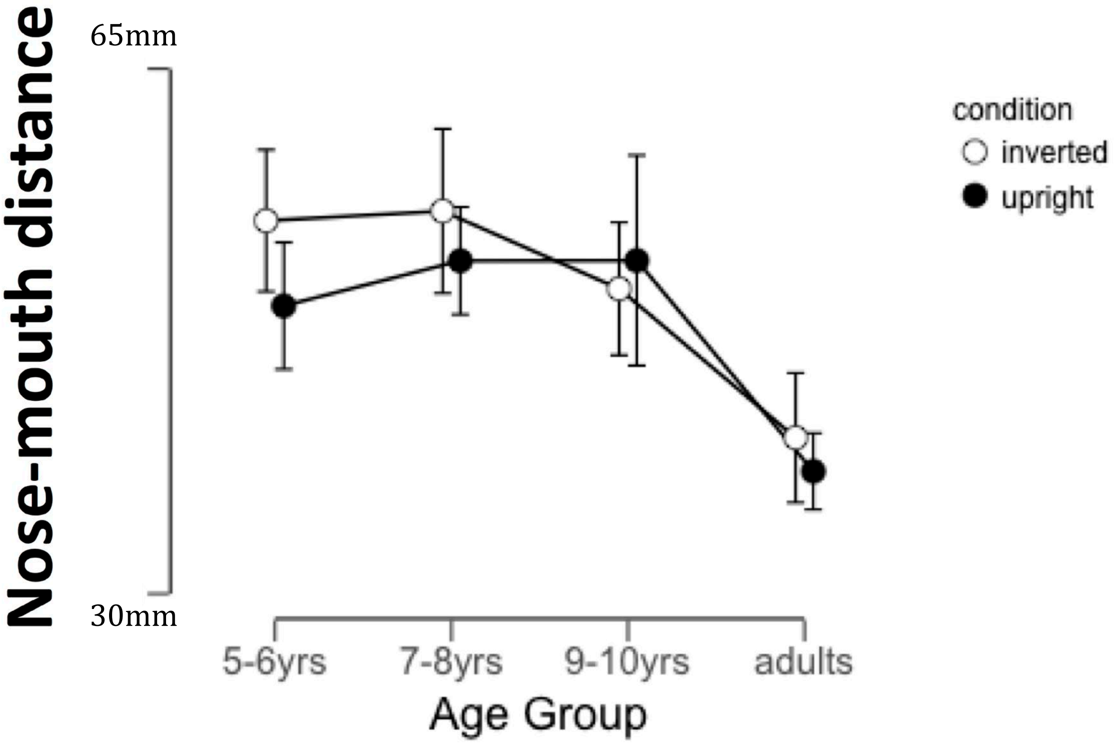

Figure 7. Average vertical distance between the nose and mouth as a function of Age Group and portrait orientation. Error bars indicate 95% credible intervals. Sample sizes for each age group are as follows: 5-6 y.o., N=24; 7-8 y.o., N=22; 9-10 y.o., N=20; Adults, N=20).

Discussion

Our results reveal several important properties of children's and adult's estimates of typical face configuration. First, across a range of different descriptors, we find that children across our entire age range (as old as 10 years of age) persist in making quantitatively larger errors than adults with regard to the typical arrangement of facial features within the facial outline. In general, this outcome is consistent with prior results describing weaker sensitivity to changes in face configuration for children as old as 10-11 years old (Mondloch et al., 2004; Mondloch, Leis and Maurer, 2002; Bauduoin et al., 2010). The positioning of the eyes too high within the facial outline, which both our child and adult participants exhibit in this task, is also consistent with Smith, Kempe and Wood's (2021) results using a similar composite task to ours. Our analysis of both the global face arrangement and specific local relationships also adds to these results by demonstrating that while there are ongoing changes related to children's vertical and horizontal placement of the eyes, there are varying effect sizes associated with errors in other spacing relationships in the face. In particular, children appear to create more expanded portraits of the face than adults, spreading the eyes, nose and mouth radially outward away from the center of the face to a greater degree. This is evident in the estimates of scaling we observed using the Procrustes distance and also in the pairwise distances we measured between facial features, each of which tends to be larger in younger children relative to adults. The high position of the eyes within the face outline after this expansion is consistent with Carbon and Wirth's (2014) "face-from-below" theory, in which they suggest that early experience viewing faces from beneath during childhood may lead to exposure to eyes high in the forehead in a foreshortened view. However, that same foreshortening should tend to compress the facial features vertically as well, which is the opposite of what we see in our data. To explain the overall expansion we observe, it could be the case that the eyes are prioritized in children's estimates of configuration and that their placement could indeed be biased by the foreshortened viewing angle described by Carbon and

Wirth. However, perhaps the remaining facial features are subsequently not matched to a global template based on biased experience, but are instead placed based on a more cognitive than perceptual approach. That is, after placing the eyes too high in the head, perhaps children reason that the nose and mouth must fill the outline more completely and place these in the outline based on this rule rather than on the basis of a perceptual template. An interesting question for further work based on this hypothesis could be to either examine which features children place in the outline first relative to the magnitude of placement errors they make or to manipulate the order in which they are permitted to place facial features to see if this changes the outcome.

Our data also indicate that there is little to no effect of inversion on the nature of the portraits constructed by children or adults. This is surprising, insofar as a number of studies including children (Baudouin et al., 2010) and adults (Leder and Bruce, 2000) have found evidence supporting the hypothesis that face inversion specifically disrupts the ability to process configural information. This interpretation of these prior results is not uncontroversial, however. In other studies, matching the difficulty of detecting featural and configural changes to faces in the upright orientation has reduced or eliminated the interaction between feature change type and face inversion (Riesenhuber et al., 2004; McKone et al. 2012). While the current study was not designed to examine a differential impact of inversion across different aspects of facial appearance, our data strongly suggest that face production is not affected by planar inversion. One potentially interesting account of this result is that it could demonstrate that errors in production result from domain-general mechanisms rather than any face-specific process. We will expand on this point in the General Discussion, but to the extent that inversion effects are a hallmark of face-specific processing, the absence of inversion effects could be interpreted as the absence of the same. Indeed, in adults, Harrison, Jones and Davies (2017) reported that drawings of houses, human faces, and cat faces each exhibited a tendency to place features (eyes or windows) higher within the object contour than a studied drawing depicted. The changes we have observed with regard to feature placement developmentally may therefore reflect the trajectory of a broader underlying bias that affects a wider range of stimuli.

Our use of a production task is useful for revealing some of the properties of children's estimates of typical face configuration relative to adults, but also requires us to be careful about the extent to which we should make inferences about *perception* based on observers' *production* of images. On one hand, the relationship between these things may be fairly direct: Drawings and other forms of image production are sometimes assumed to be faithful approximations of internal representations. This perspective motivates the application of drawing as a means of probing subjective experience for the purposes of clinical diagnosis (Pontius, 1976; Shin et al., 2006) and characterizing the experience of phenomena like visual crowding (Coates, Wagemans, and Sayim, 2017). However, there are also known dissociations between face production and face perception that suggest we should be more cautious than this. Adult portrait artists typically need to be explicitly taught rules for dividing the face into reasonable proportions (see Balas and Sinha, 2007 for an overview of such explicit instructions) despite the fact that the difference between a squashed-skull face and a typical face would be readily apparent to any naïve adult observer. Similarly, Balas and Sinha (2007) found that adult participants were generally quite poor at creating a good likeness of a familiar face using a production paradigm similar to the current one, but were far better at choosing a familiar face with the correct configuration from a set of distractors with varying arrangements of facial features. These results imply that observers make errors of production that are either larger or maybe even fundamentally different than our errors of perception. This is related to the fundamental challenge of trying to make inferences about internal representations based on observers' responses to stimuli or task demands. In the case of drawing/production tasks, it is generally the case that errors in production could be related to either the nature of the prototypes participants reference to create their drawing or the act of drawing itself. Drawings cannot therefore be interpreted uncritically as a proxy for prototypes and combining assays of perception with production tasks is a useful way to more completely characterize participants' abilities. In Experiment 2, we therefore chose to examine children's and adults' perception of the face portraits created by our participants from Experiment 1. Specifically, we investigated whether or not children in the target age range and adults could correctly identify a face matching population norms for face configuration within an array of distractors depicting the faces made by participants in each age group.

Experiment 2

In our second experiment, we examined whether children and adults could correctly identify a face with typical face configuration among foils that corresponded to the configurations created by each age group in our first task. If errors in the production of typical face geometry reflect underlying errors that affect face perception, each age group should choose the portrait made by observers from their age group as the "best" image. Alternatively, if production does not largely reflect perceptual limitations, performance in this task may be more accurate than the results from our production task.

Method

Participants. We recruited a total of 128 participants for this experiment (5-6 year-olds, N=32; 7-8 year-olds, N=32; 9-10 year-olds, N=32; Adults, N=32), none of whom took part in Experiment 1. As in our first task, participants in each group were predominantly White, with no more than 3 participants per age group self-identifying as anything other than White. Participants in this task were recruited from a wider range of locations, including participants who completed the task online due to the COVID-19 pandemic. As a result, participants' experience of adult White male faces in their environment may be more variable relative to the participants in Experiment 1. While a within-participants comparison of face production and perception would be intriguing, our intent to use the portraits from Experiment 1 in this second task made such a study design impractical. With regard to overall statistical power for our factors of interest, a post-hoc power analysis carried out in G*Power 3.1 indicates that this sample size provides us with ~80% power (1-β) to detect an effect size of 0.3. As in Experiment 1, all recruitment and testing procedures were approved by the NDSU IRB (#SM19258) in accordance with the principles outlined in the Declaration of Helsinki.

Stimuli. We created face images using the same template described in Experiment 1 that reflected the average position of the eyes, nose, and mouth in portraits made by 5-6 year-olds, 7-8 year-olds, and 9-10 year-olds. Additionally, we created a fourth image that reflected the population norm for eye, nose, and mouth position based on the data reported in Farkas (1994). We did not include an average portrait reflecting the average eye, nose, and mouth position created by adults in Experiment 1. This was partially motivated by a desire to keep the number of stimuli as small as possible to accommodate our youngest participants and also to facilitate the use of a simple symmetrical spatial layout that could be easily counterbalanced for stimulus position across participants and did not make any image markedly salient than the others by virtue of its position. As our key prediction was that children might endorse the image made by children their own age over the population norm, these four stimuli allowed us to address this issue while addressing the practical issues described above.

Procedure. We asked each participant to choose which of the four face images described above was the best approximation of typical face appearance. Specifically, we asked each participant to "Please choose which face looks the most like an ordinary person to you." Participants were given unlimited time to choose an image. The four images were arranged in a square array and printed on an 8.5 x11 sheet of paper for presentation (Figure 8). We created four versions of this array so that each face appeared in each possible location equally often across participants in each age group. For each participant, we recorded only which image they selected.

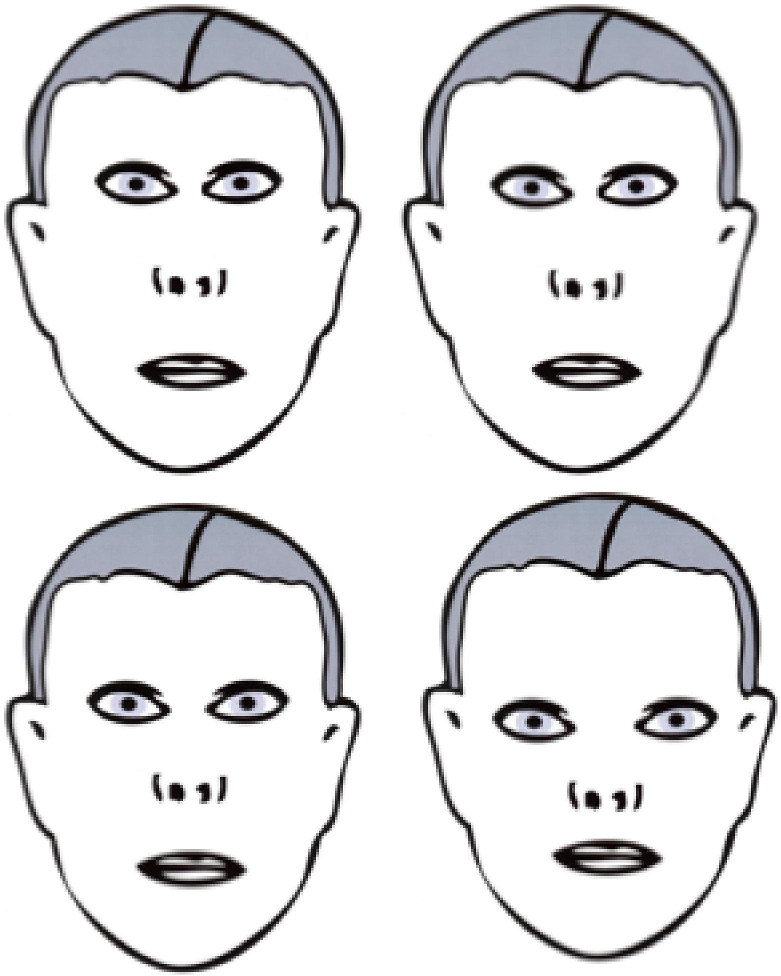

Figure 8. An example stimulus array presented to participants in Experiment 2. The upper row depicts the average configuration created by 5-6 year-olds and 7-8 year-

olds, while the bottom row depicts the average configuration made by 9-10 year-olds at left and the population norm at right. Four alternate versions of this array were used in Experiment 2 so that the position of each image was balanced across participants.

Results

In Table 1, we report the number of participants in each age group that selected each face image. We analyzed these results using a Bayesian Independent multinomial test, which revealed that the Bayes Factor for the null model of no effect of age on response proportions was extremely large (BF_{01}=58.9), indicating very strong support for the null hypothesis that age did not affect performance in this task. Moreover, each group had a clear bias favoring the face that reflected the true population average, followed by the face depicting the average face created by 9-10 year-olds and adults.

Table 1

Numbers of participants who selected each face image as representative of typical appearance. The sample size of each age group was 32 participants.

Age Group	5-6 y.o. image	7-8 y.o. image	9-10 y.o. image	Pop. Norm.
5-6 years	3	2	8	19
7-8 years	3	3	5	21
9-10 years	1	2	11	18
Adults	1	1	11	19

General Discussion

Across two experiments, we used both face production (Exp. 1) and face perception (Exp. 2) to investigate how children's understanding of face geometry develops during middle childhood. As we noted above, the simplest conclusion we can draw from our production task is that children make less accurate estimates of typical face configuration than adults, which is commensurate with many results from perceptual tasks that demonstrate children are less sensitive to specific metric relationships between facial features. Unlike these studies, however, our use of portraits allows us to comment more specifically on the nature of typical face configuration estimates as a function of age. In terms of global descriptors of face configuration, we found that children tended to make more *expanded* arrangements of features than adults, spreading the eyes, nose, and mouth further apart than is typical of adult male faces. Further, even when correcting for that expansion of the facial configuration using the Procrustes transform, we found that the residual error for children is larger than adults, suggesting that there are non-linear distortions of typical face configuration that are more pervasive in children's portraits than adults. Relative to recent data using a similar composite face paradigm (Smith, Kempe and Wood, 2021), we additionally compared performance with upright and inverted faces and found little effect of face inversion, contrary to many studies demonstrating large effects of inversion on recognition performance. This represents a new and interesting dissociation between perception and production of face geometry, one that suggests production tasks rely on different mechanisms and representations than face recognition tasks. We note, however, that this result may be task-dependent: Day and Davidenko (2018) reported an inversion effect using a time-limited face drawing task with adult observers, which may mean that inversion effects depend on drawing time or the use of a stylus rather than placement of discrete face parts as described here and in Smith, Kempe and Wood (2021).

By examining particular inter-feature relationships, we were also able to characterize the distortions evident in our tasks in more detail: Children tend to place the eyes and nose as a unit higher in the face outline than adults do, placing the mouth in more or less the right place. This leads to substantially higher nose-mouth distances in child portraits as compared to adults, and eye-nose distances that are only weakly different across age groups. This differs somewhat from prior reports describing eye-specific changes in perceptual sensitivity to vertical displacements of the eye and mouth during childhood and adolescence that were not observed for changes in mouth position (de Heering and Schilz, 2013). We suggest that our production data offers an interesting perspective on these data (and related studies) regarding the perceptual units we should consider when evaluating sensitivity to face configuration or manipulating face geometry. Specifically, our results suggest that considering the eyes and nose to be one unit may be a more appropriate choice than considering all discrete facial features to be constituent pieces of the larger face pattern, which we note is consistent with recent developmental results motivated by computational studies designed to determine which fragments of face patterns are most diagnostic (Balas et al., 2020). Our data thus reveal specific patterns of face configuration errors that persist throughout middle childhood and may indicate particular aspects of facial appearance (joint representation of the eyes and nose) that are more robust early in development. However, these errors in production are complemented by resilient perceptual abilities to identify a typical face configuration from a set of foils. Children's ability across our target age range suggests that while production of face geometry incorporates large errors in configuration at young ages, the perception of face geometry and comparison to a population norm that reflects typical variation is mature early in childhood. This result is in agreement with prior reports describing mature preferences for typical faces over faces with distorted configuration in childhood (Short et al., 2015) and also with results reporting some developmental change between early and middle childhood. Specifically, while Cooper et al. (2006) did observe that younger children (4- and 9-year-olds) did not differentiate between faces with typical feature placement and those with low feature placement in terms of attractiveness, they also reported that faces with high feature placement were indeed rated as less attractive by these children. This result indicates developmental change in the perception of attractiveness using configuration as a tool to manipulate typicality, but also demonstrates that the production errors children in our study (and in Smith, Kempe and Wood, 2021) made are not reflected in perceptual preferences.

An important consideration regarding our results is the extent to which our findings may be affected by the choice of an adult White male face as the basis for participants' portraits and the perceptual judgments we asked for in Experiment 1. Though our participant samples were predominantly White, how would our results look if we had included stimuli and analytical tools based on craniofacial data from a range of ethnicities (Ngeow and Aljunid, 2009) or from male and female faces? Children's experience with male and female faces is biased in favor of female faces early in life (Sugden et al., 2014) and children also see many faces of children near their own age. Could children have different abilities to reproduce the appropriate configuration if they were tested with stimuli that more closely matched their experience? This is an intriguing issue to investigate further and one that our data does not allow us to speak to. Our results from Experiment 1 include gross errors of face configuration, however, which do not obviously reflect the misapplication of a template based on children's experience of faces belonging to other age, race, or gender categories. While it could be the case that these errors would be reduced if we asked children to create portraits of individuals belonging to a category that they had more exposure to, we predict that children would likely persist in making these substantial errors. Smith, Kempe, and Wood (2021) examined eye placement using both male and female cartoon faces and reported no difference in eye placement as a function of participant or stimulus gender, for example. Though this is an important issue for future research, at present we think it is likely that the pattern of errors we report here may reflect general effects of children's experience viewing faces from below, or interpreting bounding contours of the face differently than adults (Carbon and Wirth, 2014).

We suggest that these results offer a useful complementary perspective on how the processing of face configuration may change during childhood, but may also be suggestive of less domain-specific changes in visual recognition mechanisms and mechanisms supporting drawing. The lack of an inversion effect in our production data is intriguing and may indicate that the phenomenology of drawing errors in face portraits is not truly a face-specific effect but indicative of more general biases in relating perceptual prototypes to graphic representations. Along these lines, we note that Harrison, Jones and Davies (2017) demonstrated similar patterns of configural errors in adult drawings of human faces, cat faces, and houses, with a clear effect of visual memory on the magnitude of these errors. The "Squashed-Skull Effect" and its development may be more like a generic "Squashed-Space Effect" that extends to a wider variety of stimuli – a possibility that should certainly be examined developmentally. Were these errors in production observed during development the result of domain-general maturation of configural information processing, this would be consistent with prior reports demonstrating that the development of sensitivity to such information may also reflect changes in domain general mechanisms for encoding appearance (Robbins et al., 2011). Also, in terms of reaching a broader understanding of the mechanisms contributing to our results, we think it is also important to consider a more general view of the face-specific mechanisms that may contribute to performance in production tasks like this above and beyond any domain-general mechanisms governing drawing. Specifically, though we have used the word "configuration" throughout this study to describe the layout of facial features within an outline, our results should not be interpreted solely within a framework that assumes observers actually calculate inter-feature distances for the purposes of face detection and recognition. Indeed, as we pointed out in the introduction, there are many good reasons to be highly skeptical of this hypothesis (Taschereau-Dumouchel, 2010; Noyes and Jenkins, 2017). Instead, we argue that our results tell us something more general about the tuning of face representations with age. Though participants manipulated segmented features individually (and we characterized them in terms of their geometric relationships to one another), the resulting portraits are estimates of global face appearance that can be used to formulate broader hypotheses about how estimates of what a typical face is change during childhood. Whether the overall appearance and layout of a face is encoded via 2nd-order metric distances (which we suggest is unlikely, for the reasons discussed previously) or via overlapping templates of intermediate-scale facial regions (Peterson, 1996), asking observers to produce images that reflect their sense of typical appearance is a valuable tool for probing internal representations, and frees us from being locked in to a particular choice of primitives for describing the images observers create for us. The dissociation we have observed between production and perception during development suggests intriguing questions for future work and we believe provides useful insights into how face layout is perceived, encoded, and recalled at different developmental stages.

References

Balas, B.J, and Sinha, P. (2007) Portraits and perception: configural information in creating and recognizing face images. *Spatial Vision, 1-2,* 119-135. DOI: 10.1163/156856807782753949.

Balas, B., Auen, A., Saville, A., Schmidt, J., and Harel, A. (2020). Children are sensitive to mutual information in intermediate-complexity face and non-face features. *Journal of vision*, *20*(5), 6. https://doi.org/10.1167/jov.20.5.6

Baudouin, J-Y., Gallay, M., Durand, K. and Robichon, F. (2010) The development of perceptual sensitivity to second-order facial relations in children. *Journal of Experimental Child Psychology, 107,* 195-206. DOI: 10.1016/j,jecp.2010.05.008.

Carbon, C-C. and Wirth, E.B. (2014) Neanderthal paintings: Production of prototypical human (Homo Sapiens) faces shows systematic distortions. *Perception, 43,* 99-102.

Coates, D.R., Wagemans, J. and Sayim, B. (2017) Diagnosing the periphery: Using the Rey-Osterrieth Complex Figure Drawing test to characterize peripheral visual function. *iPerception, 3,* 1-20. DOI: 10.1177/2041669517705447

Cooper, P.A., Geldart, S.S., Mondloch, C.J. and Maurer, D. (2006) Developmental changes in perceptions of attractiveness: a role of experience? *Developmental Science, 9,* 530-546.

Day, J. A., and Davidenko, N. (2018). Physical and perceptual accuracy of upright and inverted face drawings. *Visual Cognition, 26*(2), 89–99.

De Heering, A. and Schiltz, C. (2013) Sensitivity to spacing information increases more for the eye region than for the mouth region during childhood. *International Journal of Behavioral Development, 37,* 166-171.

Donnely, N. and Hadwin, J. (2003) Children's perception of the Thatcher Illusion: Evidence for development in configural face processing. *Visual Cognition, 10,* 1001-1017. DOI: 10.1080/13506280344000202

Edwards, B. (1999) The New drawing on the right side of the brain. Thartcher/Putnam, New York, NY.

Farkas, L.G. (1994) Anthropometry of the head and face in medicine. (2nd ed.) Elsevier, New York, NY.

Farkas, L. G., Hreczko, T. A., Katic, M. J. (1994). Craniofacial norms in North American Caucasians from birth (one year) to young adulthood. In Farkas, L. G. (Ed.), Anthropometry of the head and face (2nd ed.). (pp. 241–318) (New York: Raven Press).

Freire, A. and Lee, K. (2001) Face recognition in 4- to 7-year-olds: Processing of configural, featural, and paraphernalia Information. *Journal of Experimental Child Psychology, 80,* 347-371. DOI: 10.1006/jecp.2001.2639

Harrison, N.R., Jones, J. and Davies, S.J. (2017) Systematic distortions in vertical placement of features in drawings of faces and house. *I-Perception, 8,* 1-13.

JASP Team (2018). JASP (Version 0.8.6)[Computer software].

Kendall, D. (1989) A survey of the statistical theory of shape. *Statistical Science, 4,* 87-99. DOI: euclid.ss/1177012582

Kosslyn, S. M., Heldmeyer, K. H., and Locklear, E. P. (1977). Children's drawings as data about internal representations. *Journal of Experimental Child Psychology*, *23*, 191–211. DOI: 10.1016/0022-0965(77)90099-6

Leder H., Bruce V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*, 513–536. DOI: 10.1080/713755889

Leder H., Carbon C. C. (2006). Face-specific configural processing of relational information. British Journal of Psychology, 97, 19–29. DOI: 10.1348/000712605X54794

Lewis, M.B. (2003) Thatcher's children: development and the Thatcher Illusion. *Perception, 32,* 1415-1421. DOI: 10.1068/p5089

Long, B., Fan, J.E. and Frank, M.C. (2018) Drawings as a window into developmental changes in object representation. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

McKone, E. and Boyer, B.L. (2006) Sensitivity of 4-year-olds to featural and second-order relational changes in face distinctiveness. *Journal of Experimental Child Psychology, 94,* 134-162. DOI: 10.1016/j.jecp.2006.01.001.

McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: experience is less important than previously believed. *Cognitive neuropsychology*, *29*(1-2), 174–212. https://doi.org/10.1080/02643294.2012.660138

Mcmanus, C., Chamberlain, R., Christopherson, C., Prince-Mobbs, L., Robinson, M., and Stelk, I. (2012). Six tea-towels, one calendar and 1659 children's self-portraits: A developmental study of children drawing faces. *Perception*, 41, 235-235.

Mondloch, C.J., Dobson, K.S., Parsons, J. and Maurer, D. (2004) Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology, 89,* 159-181. DOI: 10.1016/j.jecp.2004.07.002

Mondloch, C.J., Le Grand, R. and Maurer, D. (2002) Configural face processing develops more slowly than featural face processing. *Perception, 31,* 553-566. DOI: 10.1016/j.jecp.2004.07.002

Mondloch, C.J., Leis, A. and Maurer, D. (2006) Recognizing the face of Johnny, Suzy, and me: Insensitivity to the spacing among features at 4 years of age. *Child Development, 77,* 234-243. DOI: 10.1111/j.1467-8624.2006.00867.x

Ngeow, W. C., & Aljunid, S. T. (2009). Craniofacial anthropometric norms of Malays. Singapore medical journal, 50(5), 525.

Noyes, E. and Jenkins, R. (2017) Camera-to-subject distance affects face configuration and perceived identity. *Cognition, 165,* 97-104. DOI: 10.1016/j.cognition.2017.05.012

Ostrofsky, J., Cohen, D., and Kozbelt, A. (2014). Objective versus subjective measures of face-drawing accuracy and their relations with perceptual constancies. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 486-497.

Ostrofsky, J., Kozbelt, A., Tumminia, M.J., and Cipriano, M. (2016). Why do non-artists draw the eyes too far up the head? How vertical eye-drawing errors relate to schematic knowledge, pseudoneglect, and context-based perceptual biases. *Psychology of Aesthetics, Creativity, and the Arts, 10*, 332-343.

Pellicano, E., Rhodes, G. and Peters, M. (2006) Are preschoolers sensitive to configural information in faces? *Developmental Science, 9,* 270-277. DOI: 10.1111/j.1467-7687.2006.00489.x

Peterson, M.A. (1996) Overlapping partial configurations in object memory. In M.A. Peterson and G. Rhodes (Eds.), *Perception of Faces, Objects, and Scenes.* Oxford University Press, NY. DOI: 10.1093/acprof:oso/9780195313659.003.0011

Pontius, A.A. (1976) Dyslexia and specifically distorted drawings of the face – a new subgroup with prosopagnosia-like signs. *Experientia, 32,* 1432-1435.

Riesenhuber, M., Jarudi, I., Gilad, S. and Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London, B. (Suppl.)*, 04BL0061.S1-S3. DOI: 10.1098/rsbl.2004.0216.

Robbins, R. A., Shergill, Y., Maurer, D., & Lewis, T. L. (2011). Development of sensitivity to spacing versus feature changes in pictures of houses: Evidence for slow development of a general spacing detection mechanism? *Journal of experimental child psychology*, *109*(3), 371–382. https://doi.org/10.1016/j.jecp.2011.02.004

Rouse, H., Donnelly, N., Hadwin, J.A. and Brown, T. (2004) Do children with autism perceive second-order relational features? The case of the Thatcher illusion. *The Journal of Child Psychology and Psychiatry, 45,* 1246-1257. DOI: 10.1111/j.1469-7610.2004.00317.x

Shin, M. S., Park, S. Y., Park, S. R., Seol, S. H., and Kwon, J. S. (2006). Clinical and empirical applications of the Rey-Osterrieth Complex Figure Test. *Nature Protocols, 1,* 892–899. DOI: 10.1038/nprot.2006.115.

Short, L.A., Mondloch, C.J. and Hackland, A.T. (2015) Attractiveness judgments and discrimination of mommies and grandmas: perceptual tuning for young adult faces. *Journal of Experimental Child Psychology, 129,* 1-11.

Smith, K., Kempe, V., and Wood, L. (2021). . *i-Perception*, *12*(3), 20416695211017564.

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, 56(2), 249-261.

Tanaka, J.W. and Gordon, I. (2011) Features, configuration, and holistic face processing. Oxford Handbook of Face Perception, Oxford University Press. DOI: 10.1093/oxfordhb/9780199559053.013.0010

Taschereau-Dumouchel, V., Rossion, B., Schyns, P.G. and Gosselin, F. (2010) Interattribute Distances do not represent the identity of real world faces. *Frontiers in Psychology, 1:159.* DOI: 10.3389/fpsyg.2010.00159