



## OPEN ACCESS

## EDITED BY

Shota Uono,  
University of Tsukuba, Japan

## REVIEWED BY

Masatoshi Ukezono,  
University of Human Environments, Japan  
Zhongqing Jiang,  
Liaoning Normal University, China

## \*CORRESPONDENCE

Sophia M. Thierry  
✉ st18vy@brocku.ca

RECEIVED 25 June 2024

ACCEPTED 24 April 2025

PUBLISHED 01 September 2025

## CITATION

Thierry SM, Uddenberg S, Albohn D and  
Todorov A (2025) Capturing variability in  
children's faces: an artificial, yet realistic, face  
stimulus set.  
*Front. Psychol.* 16:1454312.  
doi: 10.3389/fpsyg.2025.1454312

## COPYRIGHT

© 2025 Thierry, Uddenberg, Albohn and  
Todorov. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Capturing variability in children's faces: an artificial, yet realistic, face stimulus set

Sophia M. Thierry<sup>1\*</sup>, Stefan Uddenberg<sup>2</sup>, Daniel Albohn<sup>3</sup> and  
Alexander Todorov<sup>3</sup>

<sup>1</sup>Department of Psychology, Brock University, St. Catharines, ON, Canada, <sup>2</sup>Department of Psychology, The University of Illinois Urbana-Champaign, Champaign, IL, United States, <sup>3</sup>Booth School of Business, The University of Chicago, Chicago, IL, United States

Children's faces are underrepresented in face databases, and existing databases that do focus on children tend to have limitations in terms of the number of faces available and the diversity of ages and ethnicities represented. To improve the availability of children's faces for experimental research purposes, we created a novel face database that contains 500 artificial images of children that are diverse in terms of both age (ages 3 to 10) and ethnicity (representing 15 different racial or ethnic groups). Using deep neural networks, we produced a large collection of synthetic photographs that look like naturalistic, realistic faces of children. To assess the representativeness of the dataset, adult participants ( $N = 585$ ) judged the age, gender, ethnicity, and emotion of artificial faces selected from the set of 500 images. The images present a diverse array of artificial children's faces, offering a valuable resource for research requiring children's faces. The images and ratings are publicly available to researchers on Open Science Framework (<https://osf.io/m78r4/>).

## KEYWORDS

children, face database, artificial face generation, expressions, emotions

## Introduction

Face stimulus databases have become increasingly more diverse and comprehensive in recent years. These databases play a crucial role in advancing facial recognition technology and conducting research in various fields, including the study of identity and emotion perception (Bindemann et al., 2012; Marini et al., 2021; Marusak et al., 2015) and developmental and social psychology (Mather and Carstensen, 2005; Widen and Russell, 2010; Willis and Todorov, 2006). These databases provide researchers with a wide range of facial images, encompassing variations in lighting, pose, expression, and demographic characteristics, among other image and face properties. Although such databases are becoming larger and more diverse, children's faces are often underrepresented in these databases, making it difficult or impossible to answer many kinds of research questions involving the perception and judgment of children. This work aims to improve researchers' access to young faces by providing a diverse set of artificial child faces that vary across perceived age and ethnicity.

Traditionally, face databases have been captured in controlled conditions using standardized settings, professional models, and limited image and model variability (Beaupré et al., 2000; Belhumeur et al., 1997; Ekman and Friesen, 1976; Georgiades et al., 2001; Phillips et al., 2000). More recently, there has been a shift towards using more naturalistic and diverse datasets to better represent the real-world variability of faces. Researchers now have access to databases that provide high-resolution standardized pictures of faces of different ages, ethnicities, and emotional expressions (Bainbridge et al., 2012; Conley et al., 2018; Huang et al., 2008; Ma et al., 2015). For example, the Labeled Faces in the Wild

database contains more than 13,000 images of faces collected from the internet (Huang et al., 2008). The availability of large sets of faces makes it easier for researchers interested in using face stimuli for different purposes to obtain a diverse collection of faces.

Nonetheless, even very large databases can underrepresent several demographics, making it more challenging for researchers to access nationally or globally representative face stimuli. The inclusion of children's faces is notably limited. As images in large databases are often sourced from the internet, the images predominantly feature young adults, making it difficult to source a large subset of ethnically diverse child faces from these larger sets. Even when child faces are available (see Chandaliya and Nain, 2022), the larger face databases are usually created for training more accurate machine learning algorithms (e.g., facial recognition systems) rather than for social science research, complicating the task of curating suitable face stimuli (Huang et al., 2008; Chandaliya and Nain, 2022; Cao et al., 2018; Karkkainen and Joo, 2021; Taigman et al., 2014). Importantly, using children's faces that are sourced from the internet poses ethical and privacy concerns, as using these images would almost certainly lack both parental consent and child assent.

The child face databases currently available for research purposes unrelated to AI training include only a limited number of faces and tend to overrepresent White faces (see Table 1 for a summary of the current landscape of child face databases). For instance, the Dartmouth Database of Children's Faces is a controlled stimulus set including photographs of 80 White children between the ages of 6 and 16 posing eight distinct emotional expressions (Dalrymple et al., 2013). If researchers are interested in a narrower age range, this greatly reduces the number of unique faces available. The Child Affective Facial Expression Set (CAFE) is another controlled stimulus set of face photographs that includes 154 children of various ethnicities between the ages of two to eight, posing six distinct emotional expressions (LoBue and Thrasher, 2014). Despite including a more diverse sample of children's faces, the CAFE database predominantly features White children (50%) among the relatively small number of photographs. The limited availability of diverse face sets, particularly for children, poses significant challenges to researchers seeking to study how we see, remember, or judge faces in an inclusive manner.

The current database was constructed to provide the broader research community with a large database of children's faces that are perceived to be diverse in terms of both age (ages 3 to 10) and ethnicity (representing 15 different racial or ethnic groups). Using deep neural networks, we produced a large collection of synthetic photographs that look like naturalistic, realistic faces of children varying in age and ethnicity, all without risking the privacy or dignity of any real person. The originating model was trained on thousands of faces sourced from the internet. Only images publicly released online under permissive non-commercial licenses were used in training the originating model and further authorization for the use of these images was not required (see Karras et al., 2019). Although there is a sampling bias, in that the underlying training set of internet images oversampled White adult faces, efforts were made to ensure diverse representation in the final dataset by deliberately curating images that look like children of various ethnicities. The resulting face set comprises a wide range of perceived ethnicities, including but not limited to Black, East Asian, Hispanic or Latinx, Middle Eastern, South Asian, and White. To assess the representativeness of the dataset, adult participants ( $N = 585$ ) judged the age, gender, ethnicity, and emotion of a subset of artificial faces from the full sample of 500 images to ensure that the generated faces represented different ages and ethnicities.

## Methods

We generated 50,000 faces at random using StyleGAN2, a generative adversarial network (GAN) capable of generating highly realistic face images (Karras et al., 2019, 2020). This model was trained on over 70,000 high-quality portrait images of real faces that naturally varied along many characteristics, including age, gender, race, and expression. Face stimuli were generated using the pretrained model "stylegan2-ffhq-config-f" from NVIDIA with a truncation  $\psi = 0.75$  (for detailed information on training, hyperparameters, and various configurations, see Karras et al., 2020). Images were sampled using the  $\mathcal{W}$  latent space (as opposed to  $\mathcal{W}+$  latent space).

All of the generated faces appeared realistic but occasionally contained abnormalities, such as misshapen earrings or other distorted elements within the image. The first author manually filtered the images for what looked like faces of children between the ages of 3 and 12. Images that exhibited irregularities in the facial features were discarded. The process continued until there were 80–100 artificial faces for each of five major perceived racial categories — Black, East Asian, Hispanic or Latinx, South Asian, and White. This process countered the model's bias toward generating images of White faces. Errors in the perceived age and race/ethnicity of the images were expected, so we gathered a large sample of images and planned for further validation by other raters. Once 500 faces were collected, the first author cropped any image that had distorted features in the background. We note that some subtle abnormalities may persist in the final dataset; however, we determined these abnormalities to be unlikely to impact the overall use of the images. Only four participants from our total sample noted in the feedback section that the faces might appear AI-generated. Researchers are welcome to use the stimuli they consider subjectively appropriate, though we provide norming data for all images.

## Participants

Our final sample consisted of 585 Prolific participants from the United States (range = 18–74 years, mean = 41.82 years; [www.Prolific.co](http://www.Prolific.co)). We excluded 7 participants who did not provide a completion code and eight participants who indicated that they did not take the task seriously (a rating less than 60 on a scale ranging from 1 to 100 on seriousness). Participants identified their gender as male ( $n = 333$ ), female ( $n = 237$ ), transwoman ( $n = 4$ ), transman ( $n = 1$ ), a gender not listed ( $n = 5$ ) or preferred not to answer ( $n = 5$ ). Participants identified their race/ethnicity as White ( $n = 394$ ), Black ( $n = 76$ ), two or more races ( $n = 41$ ), Hispanic or Latinx ( $n = 31$ ), East Asian ( $n = 14$ ), Southeast Asian ( $n = 12$ ), South Asian ( $n = 8$ ), Middle Eastern ( $n = 1$ ), Native American ( $n = 1$ ), a race/ethnicity not listed ( $n = 4$ ), or preferred not to answer ( $n = 3$ ).

## Procedure

Each participant viewed a random sample of 25 unique artificial child faces from the larger set of 500 images. Five images were shown twice throughout the session to assess intra-rater reliability. Each image was rated by at least 25 participants. A post-hoc power analysis using the ICC.Sample.Size R package revealed that this sample size

TABLE 1 Summary of existing child face databases.

Database	Authors	Image characteristics	Face age, gender, and ethnicity	Number of faces	Number of child faces (ages 3 to 12)
Child Affective Facial Expression Set (CAFE)	LoBue and Thrasher (2014)	Children posing seven emotional expressions on an off-white background. Children are wearing an off-white cape to cover their clothes.	90 girls and 64 boys between the ages of two and eight. The children are White (77), Black (27), Latinx (23), East Asian (16), and South Asian (11).	154	154
The Child Emotion Facial Expression Set: A Database for Emotion Recognition in Children	Negrão et al. (2021)	Children filmed and photographed posing seven emotional expressions on a blank background. Children are wearing a white shirt.	55% girls and 45% boys between the ages of four and six. The children are of White descent (71%), African descent (24%), and Asian descent (5%).	124	124
Child Emotions Picture Set (CEPS)	Romani-Sponchiado et al. (2015)	Children posing or naturally expressing six emotional expressions.	Nine girls and nine boys between the ages of six and 11. The children are White (14), Black (3) and Indigenous (1).	18	18
Children's Spontaneous Expressions Video Database (LIRIS-CSE)	Khan et al. (2019)	Movie clips and dynamic images of children spontaneously expressing six emotional expressions with various backgrounds.	Seven girls and five boys between the ages of six and 12. The children are described as ethnically diverse.	12	12
Dartmouth Database of Children's Faces	Dalrymple et al. (2013)	Children posing eight emotional expressions from five camera angles under two lighting conditions on a black background. Children are wearing a black bib and hat to cover clothes, hair and ears.	40 female and 40 male White models between the ages of six and 16.	80	69
National Institute of Mental Health Child Emotional Faces Picture Set (NIMH-ChEFS)	Egger et al. (2011)	Children posing five emotional expressions with direct or averted gaze on a grey background.	39 female and 20 male children between the ages of 10 and 17. Information about ethnicity was not obtained, but the authors indicate that most appear White; four girls and one boy appear non-White.	59	15
Radbound Faces Database (RaFD)	Langner et al. (2010)	Adults and children posing eight emotional expressions from five camera angles and three gaze directions on a white background. Models are wearing a black shirt.	39 White adult and 10 child models (six girls, four boys).	49	10
The University of Oregon Emotional Expression Stimulus Set (DuckEES)	Giuliani et al. (2017)	Children and adolescents posing seven emotional expressions on a white background.	22 female and 15 male models between the ages of eight and 18; 89% White.	37	10

achieved a power of 1.00, with a significant level of  $\alpha = 0.05$  for calculating the Intraclass Correlation Coefficient (ICC; Zou, 2012). This sample size is also consistent with previous face dataset validation studies (e.g., Egger et al., 2011; Langner et al., 2010; Khan et al., 2019).

In each trial, participants viewed a face at the top of the screen and were asked to estimate the child's age, gender, race/ethnicity, and

emotional expression (angry, disgusted, fearful, happy, neutral, sad, surprised). There were 15 different response options available for participants to select from regarding race/ethnicity, as well as a response option provided for participants to input a race/ethnicity that was not among the predefined options. Participants were able to select from a dropdown menu additional information about each

race/ethnicity. See [S1 Appendix](#) for the list of questions participants answered to estimate each child's age, gender, race/ethnicity, and emotional expression.

At the end of the session, participants completed a demographic survey and answered questions about their task performance. For completing the 15-min study, participants received 1.77 GBP. The procedure was approved by the Institutional Review Board at the University of Chicago.

## Results

### Reliability

To measure intra-rater reliability, each participant rated five images twice. The Intraclass Correlation Coefficient (ICC) was used to assess agreement for the age estimate. The ICC value was 0.94, indicating excellent agreement, 95% CI [0.93, 0.94]. Cohen's Kappa was used to assess agreement for the categorical variables (gender, race/ethnicity, and emotion). The agreement was almost perfect for the gender estimate,  $\kappa = 0.92$ ,  $Z = 49.80$ ,  $p < 0.001$ . The agreement was substantial for the emotion estimate,  $\kappa = 0.72$ ,  $Z = 57.90$ ,  $p < 0.001$ . Moderate agreement was observed for the race/ethnicity estimate,  $\kappa = 0.58$ ,  $Z = 101$ ,  $p < 0.001$ . With all outcomes achieving statistical significance, these results demonstrate that participants were reliable in their ratings of the four estimates.

To measure inter-rater reliability, we used ICC to assess the reliability of the age estimate. The ICC value was 0.53, indicating moderate reliability, 95% CI [0.50, 0.56]. Given that not all faces were rated by the same number of raters, we used percent agreement to measure inter-rater reliability for the categorical variables. The percent agreement for gender was 90% (range = 0.5–1.00), 47% for race/ethnicity (range = 0.12–1.00), and 74% for emotion (range = 0.26–1.00). The percent agreement for estimates of race/ethnicity is lower due to the larger number of available response options, but still significantly above chance level (0.07),  $t(499) = 33.13$ ,  $d = 1.48$ ,  $p < 0.001$ .

### Subjective ratings

For each image, we calculated the mean age estimate and the number of participants who classified each image as a boy or girl, one or more of the 15 racial/ethnic categories, and one of the seven emotion categories. Then, we calculated the proportion of participants who classified each face as a boy or girl, one or more of the 15 racial/ethnic categories, and one of the seven emotion categories. Most faces are perceived as between the ages of 3 to 10 ( $n = 486$ ); there are at least 12 images of boy and girl faces that fall in each of these age groups. Very few faces were perceived as over the age of 10 ( $n = 4$ ) or under the age of three ( $n = 10$ ). Our initial aim was to select images of children between the ages of 3 to 10, but we included the additional faces that fall outside of this age range, nonetheless. [Figure 1](#) shows the number of perceived boy and girl faces in each age group. The top five dominant perceived races/ethnicities across the 500 faces are White ( $n = 195$ ), Latinx ( $n = 79$ ), Black ( $n = 77$ ), Chinese ( $n = 39$ ), and Filipinx ( $n = 26$ ). The top five dominant perceived emotions are happy ( $n = 315$ ), neutral ( $n = 151$ ), surprised ( $n = 13$ ), sad ( $n = 10$ ), and happy/neutral ( $n = 5$ ).

To better balance the distribution of faces based on age and race/ethnicity, we created a supplemental database by excluding faces where the dominant perceived race/ethnicity was White until the proportion of faces perceived as White was no greater than 20% for each age group (3 to 10 years). This resulted in a total of 392 artificial images that are perceived to fall within the age range of 3 to 10 years. [Figure 2](#) shows the distribution of perceived race/ethnicity by age group and [Figure 3](#) shows the distribution of perceived emotion by age group for both the full and reduced sample of faces. Sample images of faces perceived between the ages of 3 to 10 displaying happy and neutral expressions are shown in [Figure 4](#). Summary tables of these characteristics for the full and reduced sample can be found on our Open Science Framework page for this database.

### External validation

To further validate the ratings, we conducted an exploratory analysis to examine whether participants' subjective ratings corresponded to attribute predictions using a generative adversarial network (GAN) (see [Peterson et al., 2022](#)). The attribute prediction model was developed from adults' ratings of artificially generated faces from across the lifespan. We performed Spearman rank-order correlations to examine the relationship between our four attributes and the corresponding attributes in Peterson et al.'s model. For each image ( $n = 500$ ), a prediction score was generated from Peterson et al.'s model in standard deviations. We took the mean or proportion scores from the subjective ratings and correlated these scores to the model's prediction scores at the level of the images. We found that participants' subjective ratings correlated to the model predictions for age ( $r = 0.32$ ,  $p < 0.001$ ), gender (male:  $r = 0.61$ ,  $p < 0.001$ ), happiness ( $r = 0.61$ ,  $p < 0.001$ ) and race/ethnicity (e.g., Black:  $r = 0.63$ ,  $p < 0.001$ ; White:  $r = 0.78$ ,  $p < 0.001$ ) see [Supplementary Figure 1](#). The age correlation coefficient is likely lower than the other correlations because in our study, participants rated the age of the faces on a scale from 0 to 18, whereas the model's predictions are based on faces spanning the entire lifespan. Since our sample only included children's faces, the model predictions in SDs were all negative, reducing variability in the prediction score. For a complete table of the correlations between our 15 race/ethnicity classifications and the seven race/ethnicity classifications available in the model predictions, see [Supplementary Table 1](#).

## Discussion

The Artificial Child Face Database presents 500 artificial images of (what appear to be) children of varying ages and ethnicities. After selecting a set of 500 faces generated via a deep-learning model, we conducted a validation study to ensure that the generated faces represented different ages and ethnicities.

This database provides a valuable starting point for researchers interested in using a diverse and highly variable set of child faces for their studies. While it does not present an equal representation of all age groups and ethnicities, it does offer a substantial selection of children's faces from various perceived ethnic backgrounds, allowing researchers to select subsets of children's faces for their research goals. In this database, many faces are perceived as more than one race/ethnicity, highlighting the importance of using a less constrained



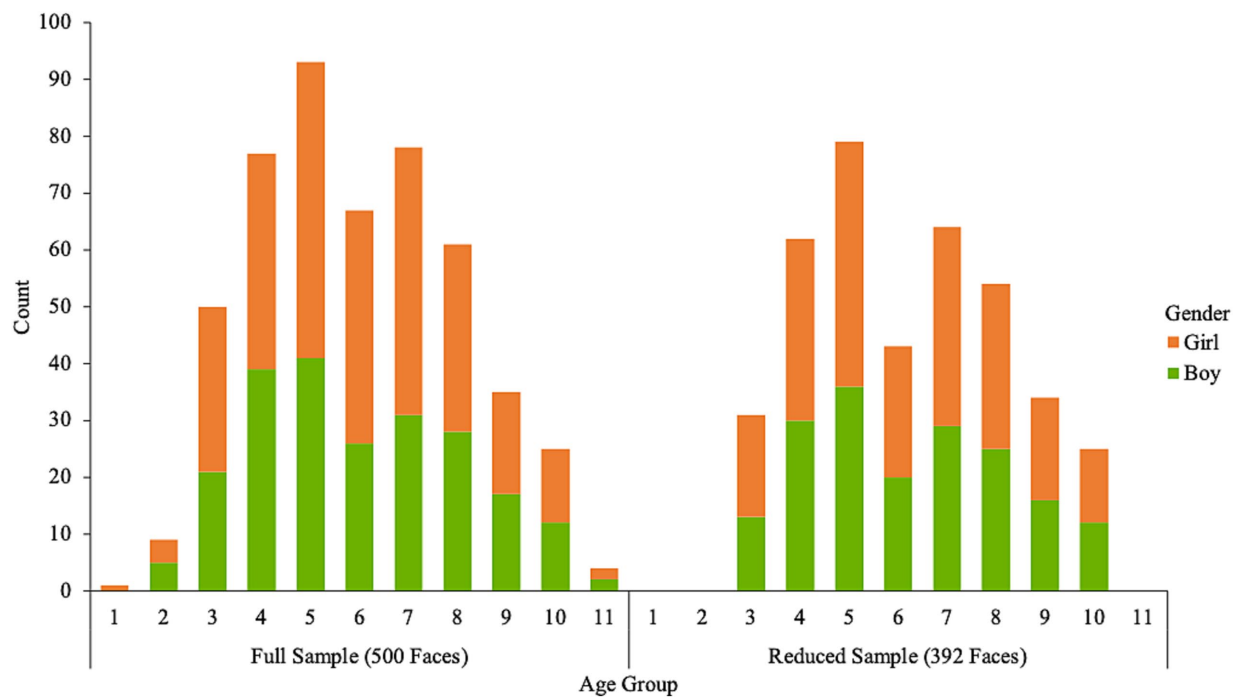


FIGURE 1  
The number of perceived artificial boy and girl faces by age group and sample.

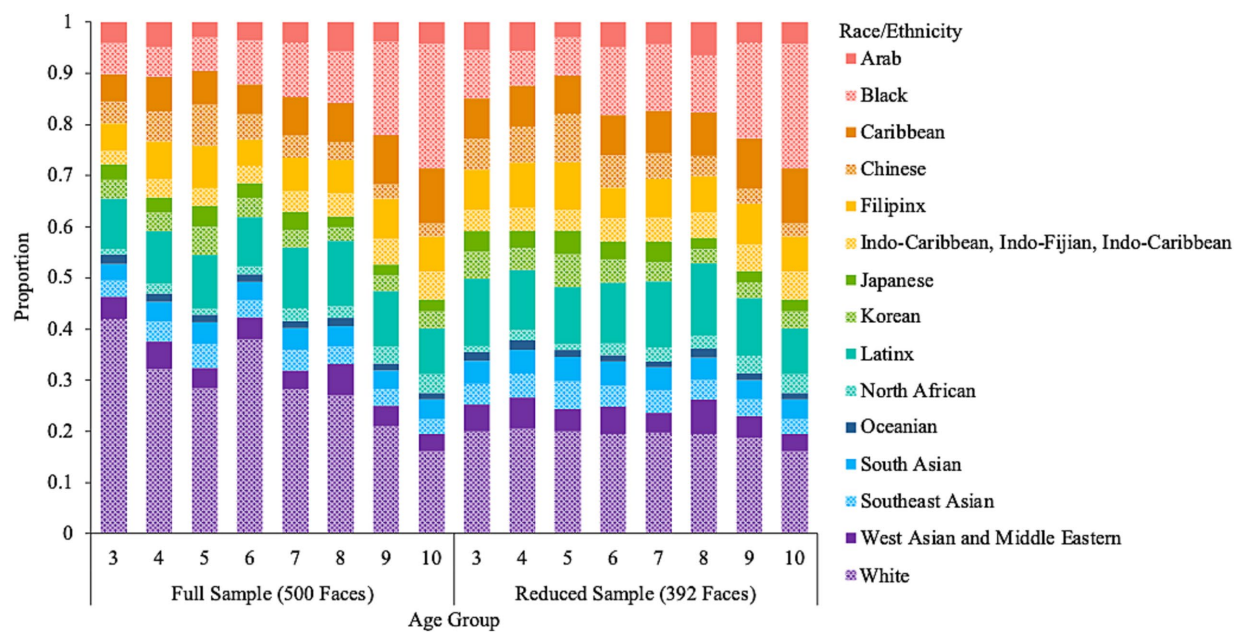


FIGURE 2  
The proportion of ethnicities perceived among the artificial faces by age group and sample.

rating measure. A forced-choice task with fewer options could have led participants to make different choices (Iankilevitch et al., 2020; Nicolas et al., 2019). For this reason, we decided to visualize the findings based on the overall proportion of perceived race/ethnicity rather than the dominant race/ethnicity, as concentrating only on the dominant race/ethnicity might not provide a comprehensive

representation of the faces. We recommend that researchers be transparent about how they select subsets of faces from this database and report the racial/ethnic proportions to ensure a more inclusive representation of the faces.

The current database provides an additional resource for researchers seeking face stimuli, addressing the need for greater

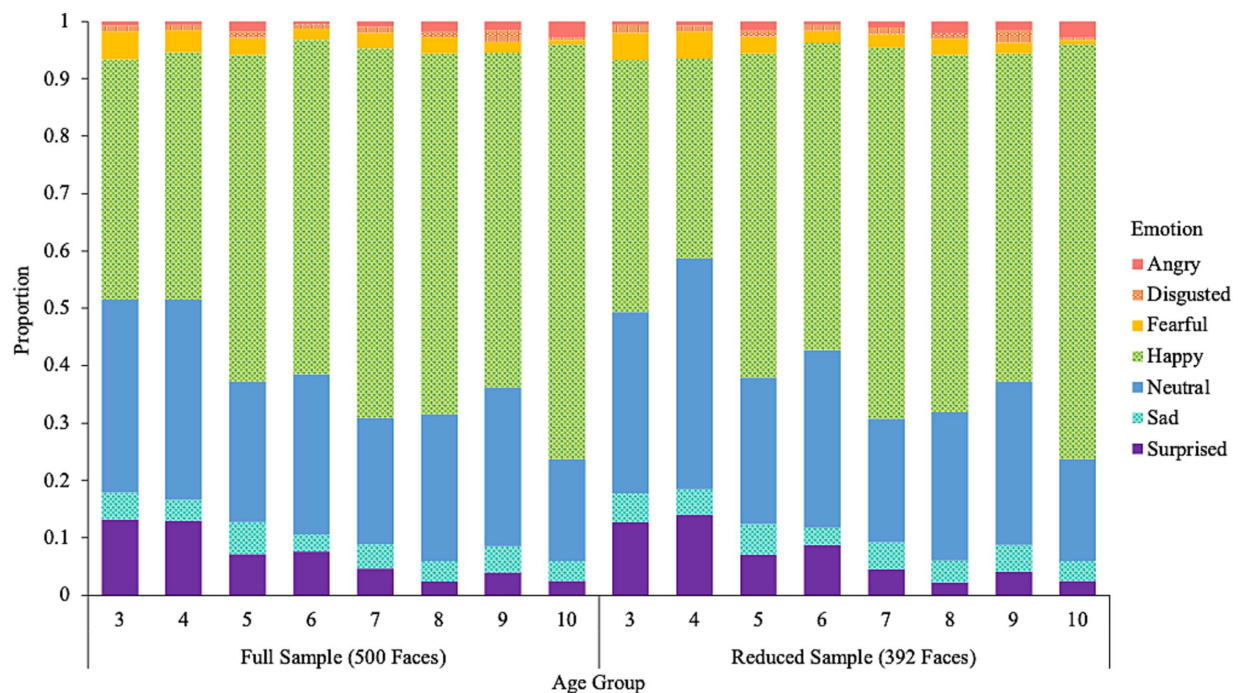


FIGURE 3

The proportion of emotional expressions perceived among the artificial faces by age group and sample.



FIGURE 4

Sample images of artificial children's faces from the artificial child face database. The faces are arranged in ascending order from perceived age three to nine (left to right), alternating girls and boys. The top row displays happy expressions; the bottom row displays neutral expressions.

inclusivity and representation of facial stimuli (Karkkainen and Joo, 2021; Cook and Over, 2021; Mondloch et al., 2023; Torrez et al., 2023). Existing child face databases mostly consist of tightly controlled images of White child faces (Dalrymple et al., 2013; Langner et al., 2010), and the databases that do include a more diverse range of images are limited in the number of faces available (LoBue and Thrasher, 2014; Egger et al., 2011; Giuliani et al., 2017; Khan et al., 2019; Negrão et al., 2021; Romani-Sponchiado et al., 2015). The limited availability of non-White children's faces restricts attempts to achieve a balanced representation of ethnicity and gender in research design. For example, our research examines first impressions of faces. Only a handful of studies have examined first impressions of children's faces. Among these studies, those that have used face stimulus sets have only used controlled images

of White children (Cogsdill and Banaji, 2015; Collova et al., 2019; Eggleston et al., 2021; Ewing et al., 2019; Talamas et al., 2016; Thierry and Mondloch, 2021). With generative AI, we saw the opportunity to create a large and highly variable database for researchers to source artificial images of children's faces. We are currently using a subset of images from this database to explore the spontaneous impressions adults and children form of children's faces varying in emotional expression, ethnicity, and age. This is a question that we would not have been able to answer without such a large collection of images.

Another advantage of using artificial faces in research is that it reduces ethical concerns associated with using real images. For instance, although face models provide informed consent for their image to be used for research, this broad consent may not fully

represent the specific ways their image will be used. Likewise, sampling faces from the internet poses additional privacy concerns, as individuals have not consented to their images being used for research purposes. These ethical concerns are particularly relevant for vulnerable populations like children who are not yet able to provide informed consent and who are not responsible for uploading their own images to the internet. These same concerns are present when using (even permissively licensed) photographs of real people sampled from the internet, whose uploaders may not realize that their images could be judged by thousands of strangers for research purposes. When using artificial faces, the resulting generated faces do not represent any specific person, ensuring individual privacy. We expect that the current dataset will be useful for research across a multitude of disciplines. For instance, in developmental psychology, these AI-generated images can be used in place of real images of children for research on peer interaction (e.g., Yazdi et al., 2020), resource allocation (e.g., Elenbaas et al., 2016), future-oriented thinking (e.g., Jerome et al., 2023), trust decisions (Grüneisen et al., 2021), and stereotyping (e.g., Shutts et al., 2016) to name a few examples.

The current database of children's faces is only the first step in possible avenues to use artificially generated child faces in research design. Face generation software presents opportunities to create custom-made faces for specific purposes. For example, an endless array of faces of any age can be generated (Karras et al., 2019, 2020). Notably, advanced modelling techniques can be used to generate faces based on perceived physical or social characteristics, such as age or attractiveness, and to transform face photographs along these dimensions (Peterson et al., 2022; Shen et al., 2022). To date, these methods have included faces from across the lifespan, but most faces have been those of young adults. Here, we found that participants' subjective ratings correlated with Peterson et al. (2022) model predictions for age, happiness, and race/ethnicity. Future research should explore how precisely the model's predictions apply to child faces and the facial features or demographic characteristics that might influence the model's accuracy. Additionally, researchers could expand on the existing models to develop models specifically for children's faces, aiming to understand the judgments adults and children form of children's faces. Generative models, such as Variational Autoencoders and diffusion models, are also suitable for generating child faces. The choice of which technique to use depends on the research goals (Vivekananthan, 2024). We used StyleGAN2 for image generation because it produces high-quality images and performs well with image manipulation (see Peterson et al., 2022). For instance, future work could generate child faces that appear to have a certain emotion and that vary according to specific attributes such as hair colour, "cuteness," or "niceness," to name just a few examples. These methods have applications for research on social impressions, emotion and identity perception, as well as social psychology more broadly.

Researchers interested in using artificially created faces should note that these faces, although realistic, might result in different conclusions than using real faces (Balas and Pacella, 2015, 2017; Miller et al., 2023; Nightingale and Farid, 2022). Although research shows that humans cannot successfully distinguish synthetic from artificially generated faces, this should be verified with our database (Nightingale and Farid, 2022; Boyd et al., 2023;

Shen et al., 2021; Tucciarelli et al., 2022). Another question for future directions could be to examine whether the perceived realism of the image or image quality varies according to perceived age, gender, or ethnicity. For instance, research shows that White AI faces are perceived as human more than human faces, highlighting the importance of considering how biased training algorithms might influence how AI faces are perceived (Miller et al., 2023; Nightingale and Farid, 2022). This database can be used to broadly increase diversity in the faces sampled for research purposes; however, it should not be used to compare faces based on perceived race/ethnicity without knowing how these AI-generated faces compare to real human faces, as this may introduce biases in racial representation.

Our aim with this database is to provide a large and diverse stimulus set of child faces for researchers to use in a variety of disciplines. Researchers interested in using these faces are encouraged to share their findings and additional ratings of these images to contribute to our Open Science Platform and help expand the database further to the research community. We hope that this database makes it more accessible for researchers to use a larger and more inclusive selection of child faces in their projects.

## Data availability statement

The datasets presented in this study can be found in online repositories. This data can be found here: <https://osf.io/m78r4/>.

## Ethics statement

The study involving humans were approved by the Social and Behavioural Sciences Institutional Review Board of the University of Chicago. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

ST: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. SU: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. DA: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. AT: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.



## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1454312/full#supplementary-material>

## References

- Bainbridge, W. Isola, P. and Blank, I. (2012). "Establishing database studying human face photograph memory." In *Proceedings Annual Meeting Cognitive Science Society*.
- Balas, B., and Pacella, J. (2015). Artificial faces are harder to remember. *Comput. Hum. Behav.* 52, 331–337. doi: 10.1016/j.chb.2015.06.018
- Balas, B., and Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Comput. Hum. Behav.* 77, 240–248. doi: 10.1016/j.chb.2017.08.045
- Beaupré, M. G., Cheung, N., and Hess, U. (2000). Montreal set facial displays emotion. Quebec, Canada: Montreal.
- Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 711–720. doi: 10.1109/34.598228
- Bindemann, M., Avetisyan, M., and Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *J. Exp. Psychol. Appl.* 18, 277–291. doi: 10.1037/a0029635
- Boyd, A., Tinsley, P., Bowyer, K., and Czajka, A. (2023). The value of AI guidance in human examination of synthetically-generated faces. *Proc AAAI Conf Artif Intell. Association for the Advancement of Artificial Intelligence (AAAI)*: 37, 5930–5938.
- Cao, Q., Shen, L., Xie, W., Parkhi, OM, and Zisserman, A. (2018). "VGGFace2: a dataset for Recognising faces across pose and age." In *2018 13th IEEE international conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE.
- Chandaliya, P. K., and Nain, N. (2022). ChildGAN: face aging and rejuvenation to find missing children. *Pattern Recogn.* 129:108761. doi: 10.1016/j.patcog.2022.108761
- Cogsdill, E. J., and Banaji, M. R. (2015). Face-trait inferences show robust child–adult agreement: evidence from three types of faces. *J. Exp. Soc. Psychol.* 60, 150–156. doi: 10.1016/j.jesp.2015.05.007
- Collova, J. R., Sutherland, C., and Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *J Pers Soc Psychol. American Psychological Association*: 117:900. doi: 10.1037/pspa0000167
- Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., et al. (2018). The racially diverse affective expression (RADIAE) face stimulus set. *Psychiatry Res.* 270, 1059–1067. doi: 10.1016/j.psychres.2018.04.066
- Cook, R., and Over, H. (2021). Why is the literature on first impressions so focused on White faces? *R Soc Open Sci* 8: 211146. doi: 10.1098/rsos.211146
- Dalrymple, K. A., Gomez, J., and Duchaine, B. (2013). The Dartmouth database of children's faces: acquisition and validation of a new face stimulus set. *PLoS One* 8:e79131. doi: 10.1371/journal.pone.0079131
- Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., et al. (2011). The NIMH child emotional faces picture set (NIMH-CHEPS): a new set of children's facial emotion stimuli. *Int. J. Methods Psychiatr. Res.* 20, 145–156. doi: 10.1002/mpr.343
- Eggleston, A., Geangu, E., Tipper, S. P., Cook, R., and Over, H. (2021). Young children learn first impressions of faces through social referencing. *Sci Rep* 11:22172. doi: 10.1038/s41598-021-94204-6
- Ekman, P., and Friesen, W. V. (1976). Measuring facial movement. *J. Nonverbal Behav.* 1, 56–75. doi: 10.1007/BF01115465
- Elenbaas, L., Rizzo, M. T., Cooley, S., and Killen, M. (2016). Rectifying social inequalities in a resource allocation task. *Cognition* 155, 176–187. doi: 10.1016/j.cognition.2016.07.002
- Ewing, L., Sutherland, C. A. M., and Willis, M. L. (2019). Children show adult-like facial appearance biases when trusting others. *Dev. Psychol.* 55, 1694–1701. doi: 10.1037/dev0000747
- Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. (2001). From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 643–660. doi: 10.1109/34.927464
- Giuliani, N. R., Flournoy, J. C., Ivie, E. J., Von Hippel, A., and Pfeifer, J. H. (2017). Presentation and validation of the DuckEES child and adolescent dynamic facial expressions stimulus set. *Int. J. Methods Psychiatr. Res.* 26:e1553. doi: 10.1002/mpr.1553
- Gruneisen, S., Rosati, A., and Warneken, F. (2021). Children show economic trust for both ingroup and outgroup partners. *Cogn. Dev.* 59:101077. doi: 10.1016/j.cogdev.2021.101077
- Huang, GB, Mattar, M, Berg, T, and Learned-Miller, E. (2008). "Labeled faces in the wild: a database for studying face recognition in unconstrained environments." in *Workshop on faces in "real-life" images: Detection, alignment, and recognition*.
- Iankilevitch, M., Cary, L. A., Remedios, J. D., and Chasteen, A. L. (2020). How do multiracial and monoracial people categorize multiracial faces? *Soc. Psychol. Personal. Sci.* 11, 688–696. doi: 10.1177/1948550619884563
- Jerome, E., Kamawar, D., Milyavskaya, M., and Atance, C. (2023). Preschoolers' saving: the role of budgeting and psychological distance on a novel token savings task. *Cogn. Dev.* 66:101315. doi: 10.1016/j.cogdev.2023.101315
- Karkkainen, K., and Joo, J. (2021). "FairFace: face attribute dataset for balanced race, gender, and age for bias measurement and mitigation." in *2021 IEEE winter conference on applications of computer vision (WACV)*. IEEE.
- Karras, T., Laine, S., and Aila, T. (2019). "A style-based generator architecture for generative adversarial networks." in *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. IEEE.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). "Analyzing and improving the image quality of StyleGAN." in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Khan, R. A., Crenn, A., Meyer, A., and Bouakaz, S. (2019). A novel database of children's spontaneous facial expressions (LIRIS-CSE). *Image Vis. Comput.* 83–84, 61–69. doi: 10.1016/j.imavis.2019.02.004
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognit. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076
- LoBue, V., and Thrasher, C. (2014). The child affective facial expression (CAFE) set: validity and reliability from untrained adults. *Front. Psychol.* 5:1532. doi: 10.3389/fpsyg.2014.01532
- Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5
- Marini, M., Ansani, A., Paglieri, F., Caruana, F., and Viola, M. (2021). The impact of facemasks on emotion recognition, trust attribution and re-identification. *Sci. Rep.* 11:5577. doi: 10.1038/s41598-021-84806-5
- Marusak, H. A., Martin, K. R., Etkin, A., and Thomason, M. E. (2015). Childhood trauma exposure disrupts the automatic regulation of emotional processing. *Neuropsychopharmacology* 40, 1250–1258. doi: 10.1038/npp.2014.311
- Mather, M., and Carstensen, L. L. (2005). Aging and motivated cognition: the positivity effect in attention and memory. *Trends Cogn. Sci. Elsevier BV*: 9, 496–502. doi: 10.1016/j.tics.2005.08.005
- Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., and Dawel, A. (2023). AI hyperrealism: why AI faces are perceived as more real than human ones. *Psychol. Sci.* 34, 1390–1403. doi: 10.1177/09567976231207095
- Mondloch, C. J., Twele, A. C., and Thierry, S. M. (2023). We need to move beyond rating scales, white faces and adult perceivers: Invited Commentary on Sutherland \ u0026amp; Young (2022), understanding trait impressions from faces. *Br J Psychol. Wiley*: 114, 504–507. doi: 10.1111/bjop.12619
- Negrão, J. G., Osorio, A. A. C., Siciliano, R. F., Lederman, V. R. G., Kozasa, E. H., D'Antino, M. E. F., et al. (2021). The child emotion facial expression set: a database for



emotion recognition in children. *Front. Psychol.* 12:666245. doi: 10.3389/fpsyg.2021.666245

Nicolas, G., Skinner, A. L., and Dickter, C. L. (2019). Other than the sum: Hispanic and middle eastern categorizations of black-white mixed-race faces. *Soc. Psychol. Personal. Sci.* 10, 532–541. doi: 10.1177/1948550618769591

Nightingale, S. J., and Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proc. Natl. Acad. Sci. USA* 119:e2120481119. doi: 10.1073/pnas.2120481119

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., and Suchow, J. W. (2022). Deep models of superficial face judgments. *Proc. Natl. Acad. Sci. USA* 119:e2115228119. doi: 10.1073/pnas.2115228119

Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1090–1104. doi: 10.1109/34.879790

Romani-Sponchiado, A., Sanvicente-Vieira, B., Mottin, C., Hertzog-Fonini, D., and Arteche, A. (2015). Child emotions picture set (CEPS): development of a database of children's emotional expressions. *Psychol. Neurosci.* 8, 467–478. doi: 10.1037/h0101430

Shen, B., RichardWebster, B., O'Toole, A., Bowyer, K., and Scheirer, WJ. (2021). "A study of the human perception of synthetic faces." in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE.

Shen, Y., Yang, C., Tang, X., and Zhou, B. (2022). InterFaceGAN: interpreting the disentangled face representation learned by GANs. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 2004–2018. doi: 10.1109/TPAMI.2020.3034267

Shutts, K., Brey, E. L., Dornbusch, L. A., Slywotzky, N., and Olson, K. R. (2016). Children use wealth cues to evaluate others. *PLoS One* 11:e0149360. doi: 10.1371/journal.pone.0149360

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). "DeepFace: closing the gap to human-level performance in face verification". in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

Talamas, S. N., Mavor, K. I., Axelsson, J., Sundelin, T., and Perrett, D. I. (2016). Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness. *J. Exp. Psychol. Gen.* 145, 603–620. doi: 10.1037/xge0000152

Thierry, S. M., and Mondloch, C. J. (2021). First impressions of child faces: facial trustworthiness influences adults' interpretations of children's behavior in ambiguous situations. *J. Exp. Child Psychol.* 208:105153. doi: 10.1016/j.jecp.2021.105153

Torrez, B., Hudson, S.-K. T. J., and Dupree, C. H. (2023). Racial equity in social psychological science: A guide for scholars, institutions, and the field. *Soc Personal Psychol Compass* 17. doi: 10.1111/spc3.12720

Tucciarelli, R., Vehar, N., Chandaria, S., and Tsakiris, M. (2022). On the realness of people who do not exist: the social processing of artificial faces. *iScience* 25:105441. doi: 10.1016/j.isci.2022.105441

Vivekananthan, S. (2024). Comparative analysis of generative models: enhancing image synthesis with VAEs, GANs, and stable diffusion [internet]. arXiv [cs.CV]. Available online at: <http://arxiv.org/abs/2408.08751> (Accessed March 1, 2025).

Widen, S. C., and Russell, J. A. (2010). Children's scripts for social emotions: causes and consequences are more central than are facial expressions. *Br. J. Dev. Psychol.* 28, 565–581. doi: 10.1348/026151009X457550d

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Yazdi, H., Heyman, G. D., and Barner, D. (2020). Children sensitive reputation when giving both Ingroup Uutgroup members. *J. Exp. Child Psychol.* 194:104814. doi: 10.1016/j.jecp.2020.104814

Zou, G. Y. (2012). Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat. Med.* 31, 3972–3981. doi: 10.1002/sim.5466