Published in final edited form as:

Vis cogn. 2013 January 1; 21(9-10): 1138–1164. doi:10.1080/13506285.2013.800622.

Developing Race Categories in Infancy via Bayesian Face Recognition

Benjamin Balas

Department of Psychology, North Dakota State University, Fargo, ND 58102

Abstract

The other-race effect emerges during infancy following the perceptual narrowing of face recognition. Other-race faces that were previously discriminable in early infancy cannot be distinguished by older infants. Presently, I discuss a Bayesian model of this process that posits that the other-race effect may be a consequence of learning to distinguish between intra-personal variation (changes to face appearance that preserve identity) and extra-personal variation (changes that do not preserve identity) in a visual environment in which a subset of race categories dominate. I demonstrate that race categories, which I have previously argued are a critical precursor to the emergence of the other-race effect in infancy, are a natural by-product of this model. Perceptual narrowing for race may thus be a natural consequence of visual experience and the estimation of face variability based on a growing number of exemplars. I describe the basic architecture of the model, its applicability to a range of visual learning scenarios, and identify critical choices one faces in applying the model to a specific perceptual task. Despite the success of the model in accounting for these behavioral results, I conclude by identifying important shortcomings of the model and describe important challenges for future efforts to characterize the development of the other-race effect computationally.

Keywords

Perceptual Narrowing; Other-race effect; Computational Modeling; Category Perception

1. Introduction

The other-race effect is a robust phenomenon in face recognition that depends critically on experience with faces. While it is well known that neonates and young infants prefer to look at faces more than other stimuli (Goren, Sarty & Wu, 1975), this initial bias appears to depend on broad perceptual biases (Simion et al., 2002; Balas, 2010) and does not reflect specialization of face processing. Only in the first year of life do infants begin to exhibit behaviors consistent with the other-race effect. Visual preferences for faces of their own race emerge relatively early (Kelly et al., 2005), while race-specific differences in face discrimination develop later (Kelly et al., 2007). The development of the other-race effect as indexed by reduced abilities to discriminate between other-race faces has been described as an instance of *perceptual narrowing* (Nelson, 2001). In this case, and other domains (Scott, Pascalis, & Nelson, 2007; Kelly et al., 2009), perceptual narrowing is understood to mean that an initial state in which a range of sensory inputs can be distinguished accurately gives way to a state in which only a subset of those patterns can be effectively discriminated. In the case of the other-race effect, this means that while faces belonging to multiple racial and ethnic groups can be discriminated early in the first year of life, by the age of 9-12 months infants are less able to discriminate between face exemplars that are not representative of their visual experience. Presently, my goal is to discuss and expand upon a computational model of perceptual narrowing in the context of the development of the other-race effect in infancy. Specifically, I will introduce and expand upon a Bayesian face recognition model

that I have previously used to examine the computational basis of perceptual narrowing for face perception in the first year of life. I describe prior models of the other-race effect and identify several factors that complicate our ability to draw conclusions from these results regarding the basis of the other-race effect in infancy. Next, I describe the basic architecture of the Bayesian model I have used in my own simulations, with an emphasis on how it differs from previous models of the other-race effect in terms of the use of difference images to encode face variability. I also identify several choices the modeler faces in applying the model to a specific perceptual task and discuss how these choices may relate to existing developmental results. Finally, I use the model to simulate infant performance in a race categorization task and demonstrate that race categorization appears to emerge naturally from the model as a function of increased exposure to faces. I close by suggesting that this model offers a particularly useful framework for understanding the mechanisms that support the development of the other-race effect and identifying interesting topics for empirical study.

Face-space and models of the other-race effect

There have been few efforts to use computational models to account for the development of the other-race effect, though there are multiple descriptions of how various computational mechanisms reflect different aspects of own- and other-race face perception. In general, computational models of the other-race effect rely on representations of facial appearance in a “face space,” (Valentine, 1991) which is a geometric framework for describing individual faces based on their projection onto a set of axes that capture perceptual variability along multiple dimensions. The dimensions of face space are assumed to capture appearance variability across the population of faces observed by an individual, and thus may differ from person to person as the visual system’s representation of face variability is adjusted in response to the environment. Valentine (1992) proposed that this experience-dependent adjustment of the axes of face space may be the basis of the other-race effect – by adjusting the axes of face space to capture the observed variability in one racial group, the variability in other-race faces may be reduced. The core hypothesis is that other-race faces may be less distinguishable from one another because too much information is lost once they are projected into an inappropriate low-dimensional space. The computational convenience of using a compact description of the faces an observer sees most often may thus be accompanied by the inconvenience of being forced to use a description of face appearance that is ineffective for some subsets of faces.

This principle – that models trained primarily with own-race faces inadequately capture other-race face variability – has been explored in several computational studies. In nearly all cases, these models use real images of human faces to define a computational face space and test various aspects of own- and other-race face representation that may be useful proxies of human face perception. A common way to define the axes of face space using a set of photographic images is via Principal Components Analysis, or PCA. Given data, like a set of face images, that is defined in terms of one set of coordinates (for images, these are initially pixel intensities), PCA identifies a rigid rotation of the coordinate axes such that variance across a population of faces is maximal along the first dimension, and each successive dimension captures the next largest chunk of variance. New images can then be represented compactly by projecting them onto some small(er) number of principal components, and using the coordinates of faces in this lower-dimensional space to describe appearance. Each face is ultimately a point plotted in a space of relatively small dimensions, and groups of faces are clouds of points in this space. The similarity between faces and/or groups of faces can be estimated via any number of distance metrics. The first few dimensions obtained via PCA often capture a great deal of the variance in face images (Sirovich & Kirby, 1987; though see O’Toole et al. (1993) for evidence that higher dimensions carry information for

identity), making representation by so-called “eigenfaces” (Turk & Pentland, 1991) a useful, data-driven strategy for computational investigations of the other-race effect. Indeed, several different computational results suggest that the projection of other-race faces onto eigenfaces obtained with own-race faces has important consequences. Other-race faces are known to be more poorly reconstructed than own-race faces following projection onto own-race eigenfaces (O’Toole et al., 1991; O’Toole et al., 1994), meaning that the axes best describing face variability in one racial group indeed differ substantially from the axes that are best for another group. Embedding high-dimensional data in a low-dimensional space via PCA is typically lossy – the original image cannot be perfectly reconstructed from the coefficients that result from projecting that image onto a small set of eigenfaces. That other-race faces have larger reconstruction error suggests that they do not “lie flat” in face space as much as own-race faces do, and thus important features that are diagnostic of identity may be lost once they are projected into the space. Further, the distribution of own-race and other-race faces within computational face spaces has several interesting properties. Other-race faces are outliers relative to own-race faces in face space (Haque & Cottrell, 2004), which may account for both the typical impairments in recognizing other-race faces and the typical advantages for categorizing them according to race (Levin, 2000; Ge et al., 2009). Outliers need not be hard to discriminate from one another however, so long as the distance between one outlier and the nearest points is large. However, other-race faces are also known to cluster more closely than own-race faces in low-dimensional face space (Caldara & Abdi, 2006), suggesting that the compression of appearance variability that Valentine proposed is indeed a computational consequence of learning eigenfaces from racially-biased input data. This compression has also been tested empirically by using behavioral data from same-different tasks to construct face-space using multidimensional scaling (Papesh & Goldinger, 2010).

What is the computational basis of the other-race effect?

These results are intriguing, and suggest that data-driven representations of facial appearance variability may be largely responsible for the adult other-race effect. However, there are several reasons why these results also complicate our ability to come up with a computational theory of the development of the other-race effect. In particular, the way the other-race effect is operationalized is not consistent across these studies. In some cases, reconstruction error is used as a proxy for other-race recognition, elsewhere the average distance between data points in face space is the variable of choice. Information theory measures are employed in some cases, while simple pattern discriminability is used in others. Overall, these studies strongly suggest that there are many aspects of appearance representation in face space that differ for other-race faces, but the multiplicity of computational analyses across different modeling studies makes it difficult to understand exactly what might be changing as a function of development. Also, the relationship between these various formal analyses of the other-race effect and the tasks that are used to test infants is not clear. Infants are typically tested using a fairly constrained set of methodologies, nearly all of which use visual preference as the measure of the other-race effect. Ideally, a computational model of the development of the other-race effect should make easy contact with these tasks so that we may compare infant and model performance directly.

Furl et al. (2002) carried out an extensive analysis of how the other-race effect manifests in different face recognition algorithms that addresses some of these criticisms, and also provides important insight into the computational basis of the ORE. Briefly, Furl et al. compared 13 different face recognition algorithms, spanning a range of different appearance representations (e.g. pixels, Gabors) and decision rules for face recognition. This investigation thus made it possible to determine how different vocabularies of features and

different techniques for training models to encode and recognize faces each contribute to the other-race effect computationally. Compared to the diversity of measures described above, one particular strength of the Furl et al. tests is that each model's performance was quantified in terms of recognition accuracy rather than some other proxy for face recognition capabilities. This facilitated direct comparisons between models, since the output of each model was matched. An important conclusion of the Furl et al. analysis was that the comparison of multiple models revealed that the other-race effect does not obtain robustly in all cases. Specifically, models that explicitly make comparisons between individuals via discriminant functions (or other direct comparisons of identity) exhibited a stronger ORE than models that did not. To expand upon this briefly, the key observation was that models that did not "warp" the face space to accommodate the observed variability across faces in the training set did not tend to show an other-race effect. Mere experience with a set of faces (accumulating examples in a face space) does not appear to be sufficient computationally for the other-race effect to emerge – rather, it is the adjustment of that space according to the demands of individuation that is crucial. The effective modeling of the other-race effect thus depends upon implementing a model that implements discrimination between faces at either the representational stage or at a later stage.

A Bayesian model of the other-race effect

To explore the computational basis of perceptual narrowing for own- and other-race face recognition during infancy, I chose to use one of the models that Furl et al. examined as the basis for simulations of infant face recognition. Specifically, I chose to adapt Moghaddam's algorithm (2000) for estimating intra- and extra-personal variation as the foundation of my own model. There are several reasons for this choice. First, the algorithm is highly successful in the context of computer vision suggesting that it does in fact capture important features of appearance for recognition. As I discuss in more detail below, the estimation of intra- and extra-personal variation by the model also relates very easily to infant face recognition paradigms. Finally, manipulating developmental experience in the context of most face-space models is fairly straightforward – learning the dimensions of face-space and the distribution of faces within it can be carried out with either a small or a larger number of samples, providing a model of how a sparse or densely populated face space impacts performance (Humphreys & Johnson, 2007). Again, however, I argue that the use of appearance variation as the basis for the model facilitates the manipulation of some particularly interesting aspects of experience. Specifically, the category structure of face perception is a variable that I chose to manipulate in my own simulations that is easily operationalized in this model and reveals some interesting properties of the other-race effect. I continue by offering a description of the model, including a discussion of the free parameters one must consider when working with real images, and briefly review my previous results describing how race categories appear to play a role in the development of the other-race effect in infancy.

2. Bayesian Face Recognition – An Overview of the Model

The foundation of the model I discuss here is the estimation of two different kinds of appearance variability: differences in appearance that preserve identity or *intra-personal variation*, and differences in appearance that do not preserve identity or *extra-personal variation* (Figure 1). Compared to other algorithms, the goal of this model is to distinguish between these two types of variability rather than to assign an individual label to each image. The rationale behind this approach is that we can easily assign an individual label to a face if we can determine whether the difference between some labeled instance and a new image is an example of intra-personal variability. If so, then we should assign the same label

to the new image – otherwise, we should assume that the new image depicts a new individual.

Both intra-personal and extra-personal appearance variability are implemented in the model via difference images, which are computed by subtracting two images from one another pixel by pixel (Figure 2). We can generate an instance of intra-personal variation by subtracting two images of the same person from one another (e.g. One image depicting a person smiling and another depicting the same person with a neutral expression). Likewise, we can generate an instance of extra-personal variation by subtracting two images of different people from one another (Figure 2). The model is subsequently trained with these differences between images rather than with original face images. One important advantage of this model is that the use of these difference images reduces the categorization problem to a 2-way comparison rather than an N-way comparison between multiple discrete categories (where N may be the number of individuals we want to be able to recognize, for example).

Using a set of difference images as training data, the model is implemented by estimating probability densities describing how intra- and extra-personal differences are distributed. By using probability densities, we make it possible to describe a new difference image in terms of how likely it is to have arisen from the intra- or extra-personal distributions, and combine the information from these distributions optimally via Bayes' rule. In general, Bayes' rule describes how to combine prior information with observed data to draw a conclusion about some parameter you care about. In our case, the parameter we care about is the category membership of difference images (are they intra- or extra-personal variation?) and the training data gives us a way to make inferences about new data points. With regard to the other-race effect, the underlying logic of this model is that the variability (difference images) observed within one racial group may differ from that observed within another racial group, leading to more accurate conclusions when new images are from the same race as the training data. I continue by describing the key steps in implementing the model using training data.

Constructing face spaces using image differences

To continue, the model must be trained with labeled instances of both intra- and extra-personal variability, which assumes some means of determining which difference images belong to which category (I will discuss the implications of this assumption later). New data, in the form of new difference images, can then be assigned to either category based on some distance metric. Here, I will assume that we use the labeled intra- and extra-personal data to estimate probability distributions in two eigenspaces, one for each kind of appearance variation. We do this by first embedding the training data for each category in a low-dimensional space using Principal Components Analysis. This is similar to previously described PCA-based models of the other-race effect, but it must be noted that PCA is being used here to find axes that describe the variance in sets of image differences rather than in sets of images. We also perform PCA separately for the intra- and extra-personal difference images. An important detail about this procedure is that we will assume that each instance of appearance variation actually contributes *two* data points to the training data, since we may subtract two parent images from one another in either order (A-B, or B-A). This means that our training data will always have a mean of zero, since the average of A-B and B-A is always zero. Also, the cloud of points defined by the set of difference images will be symmetrical – each single point will have a mirror image on the other side of the origin.

Estimating probability distributions for classification

After we project the training data from each category into a low-dimensional space using PCA, we estimate a probability density function so that we can describe the points in this

space with a small number of parameters. I will assume here that we will use a multivariate normal (MVN) distribution, but there is no reason other distributions can't be used. Using this distribution, the set of difference images in each face space will be described in terms of a normal distribution with a k -dimensional mean (μ) and a covariance matrix (Σ) that we estimate from the training data (Eqn 1.). This distribution then offers a way to quantify the likelihood of a difference image (labeled ' Δ ' here) given the parameters of a particular distribution (denoted here as Ω) including the mean, covariance, and number of dimensions (k) – in other words, how likely was this particular difference image given the difference images the model has already seen? We estimate separate distributions for the intra-personal training data and the extra-personal training data so new points can be evaluated with respect to both.

$$p(\Delta|\Omega) = \frac{1}{(2\pi)^{k/2} |\Sigma_\Omega|^{1/2}} \exp\left(-\frac{1}{2}(\Delta - \mu_\Omega)' \Sigma^{-1} (\Delta - \mu_\Omega)\right) \quad (1)$$

Finally, with these two distributions in hand (and associated parameters Ω_E and Ω_I for extra- and intra-personal, respectively) a new difference image can be classified as intra- or extra-personal variation by projecting it into each space, estimating the likelihood of observing that difference image given membership in the corresponding category, and making a Bayesian decision about which category to assign the difference image to (Eqns. 2-3). For each new difference image, we thus ultimately obtain the probability of the category (extra-personal variation, for example, denoted Ω_E here) given that particular difference image. This involves combining the likelihood described above with the category prior ($p(\Omega)$) to get the posterior probability ($p(\Omega|\Delta)$). The category prior is an *a priori* estimate of how likely each category is in the absence of data. Equation 2 assumes that we may have non-uniform category priors, while Equation 3 assumes that these are uniform and may thus be factored out. In practice, I suggest that we don't know meaningful values to assign to the prior probability of intra- and extra-personal variation, which suggests that the use of uniform priors is prudent.

$$p(\Omega_E|\Delta) = \frac{p(\Omega_E)p(\Delta|\Omega_E)}{p(\Omega_E)p(\Delta|\Omega_E) + p(\Omega_I)p(\Delta|\Omega_I)} \quad (2)$$

$$p(\Omega_E|\Delta) = \frac{p(\Delta|\Omega_E)}{p(\Delta|\Omega_E) + p(\Delta|\Omega_I)} \quad (3)$$

To summarize, we are using difference images as a means of describing how faces can differ from one another when identity is preserved compared to when it changes. We use PCA to determine a low-dimensional description of two sets of difference images, and then estimate probability densities from the data. These probability densities make it possible to describe new data in terms of likelihoods, and use Bayes' rule to make a decision about whether new data is intra- or extra-personal variation. In terms of how the model is trained, the distinction between intra- and extra-personal variation is realized at the very first step – we construct many difference images and label them appropriately for the model. The subsequent steps of projecting them into a face space and estimating probability densities are primarily tools for summarizing the data compactly and in terms of probabilities. This latter step allows us to use Bayes' rule, even in the absence of category priors for our two kinds of appearance variation.

Using the model to study development

For computer vision applications, we might be primarily concerned with actually recognizing faces by assigning an individual label to them or identifying them as familiar or unfamiliar. In each case, the most straightforward way to proceed is to define some threshold value of $p(\Omega_{\text{extra}}|\Delta)$ (which I will abbreviate as p_{extra} throughout the text) to accept or reject difference images. A threshold value of 0.5 is an obvious choice for rejecting faces as either familiar or unfamiliar, and if we wish to match a new face to one of several stored templates, looking for the difference image that provides the maximal value of p_{extra} gives us an optimal guess as to the identity of the new face. In the context of modeling the developing visual system, however, I suggest that the absolute value of p_{extra} is of less relevance than the relative values of p_{extra} for paired test stimuli. Infant other-race tasks often make use of the visual-paired comparison (VPC) paradigm (Fantz, 1964; Pascalis, de Haan, & Nelson, 2002; Kelly et al., 2007). In these tasks, infants are familiarized with a single face for some amount of time (10-20 seconds), after which they are presented with two test stimuli: An image that matches the familiarization stimulus and a novel image. Discrimination ability is measured by calculating the proportion of time the infant spends looking at the novel image. In this setting (and other related tasks), we can describe the infant's task computationally in terms of a comparison between p_{extra} values. If the difference image involving the novel test stimulus reliably yields a larger probability of extra-personal variation than the familiar test stimulus, the infant (and the model) should prefer that image (Figure 3). Both probabilities may be low, especially if the model has been trained on a fairly small number of images, however this is immaterial to predicting the bias for one stimulus compared to the other.

Since my goal is to explain why this model is a particularly useful means of understanding the development of the other-race effect computationally, it's critically important to acknowledge what decisions must be made when working with real data. Some of these decisions present challenges to using the model in a meaningful way, while others present opportunities to make predictions about how development proceeds. In general, I do not wish to argue that this approach is a panacea, and demonstrating that this model (or any model) "works" is not the ultimate goal. Many different approaches may work – the question is whether or not the manner in which they work can provide insight into the phenomenon we want to understand more clearly. Also, it is important to acknowledge what aspects of the model require us to make assumptions that are either not clearly motivated by known aspects of visual development, or do not match up with what is known.

What representation of the images should we use?

There are several decisions we must make as we construct the intra- and extra-personal face spaces that will serve as the basis for probabilistic inference in the model. First, the model as I have described it relies upon pixel-by-pixel differences between two images to capture appearance variation, but there are many other representations that could be used: The training and test images could be manipulated to reflect constraints on the developing visual system, for example (Kellman & Arterberry, 2007), or non-linearities that reflect the computations that are implemented in the early visual system could also be incorporated (Freedman & Dannemiller, 1996). Faces could also be described using summary statistics to capture variation in texture and pigmentation (Hancock, Burton, & Bruce, 1996) – the latter may be useful for including explicit measurement of facial features like freckles, that are likely to vary in terms of absolute location in the image, but be relatively consistent in terms of a texture description. At present we do not have principled reasons for implementing one representation of appearance over another, but identifying what features are available to infants during the first year of life may be an important way in which models like the current one can be improved upon.

How much training data?

Second, once we have chosen a representation of image appearance that we are satisfied with, we must decide how many training images of each type we wish to include. If we have

a database of n individuals with k images of each individual, there are $n \binom{k}{2}$ and $k^2 \binom{n}{2}$ potential intra- and extra-personal difference images that we could conceivably construct and use as training data. These numbers are quite large for even relatively small numbers of n and k , meaning that we must decide how we want to sample differences images from the set of possible options. If $n=100$ and $k=10$, for example, there are 4500 intra-personal difference images and approximately 500,000 extra-personal difference images. It is important to note that the set of extra-personal differences typically will be much larger than the set of intra-personal differences – whether the training data should reflect this, and how, is an interesting issue to consider, especially when we consider where labeled data is likely to come from in the visual environment (an issue I take up in the discussion).

How should we classify new difference images?

Finally, once we have a set of training data and have embedded those data points in a low-dimensional space using PCA, we must choose how to describe the distribution of intra- and extra-personal variation. In practice, a multivariate normal distribution centered at the origin is a convenient choice for both distributions. Moghaddam (2000) observed that a large sample of training data was well described by such a distribution, and that the distribution of extra-personal variation is rotated slightly relative to the distribution of intra-personal variation. This suggests that we should use a full covariance matrix to describe the shape of the MVN distributions we estimate using the training data, but our ability to do so is potentially limited by the number of data points we wish to include in the training data and the dimensionality of the face space in which we embed those points. However, note that the ability to use Bayes' Law to make a decision about the kind of variability does not depend upon any particular distribution. One could just as easily fit any other probability density function (e.g. a mixture-of-Gaussians) to the training data and use that to estimate the likelihoods that are needed to compute the posterior probabilities in each case. In any case, however, we must be mindful of the trade-offs between the dimensionality of the data, the amount of training data, and the parameters of the probability distributions we wish to estimate.

3. Race Categories and the Development of the ORE

Using the model as described above, I previously demonstrated that the developmental trajectory observed in perceptual narrowing (early competence for both races, late competence for own-race face discrimination only) was evident if race categories were incorporated into the model's training data in a particular way (Balas, 2012). Specifically, I showed that early competence with own- and other-race faces appeared to depend critically on the inclusion of difference images across race categories (e.g. A difference image created by subtracting an image of an Asian face from an image of a White face). When these comparisons contributed to the training data, I found that the model showed reliable above-chance performance in simulated VPC trials of both races. When these comparisons were not permitted, the model only performed at above chance levels for own-race faces. I suggested, consistent with recent empirical results (Anzures et al., 2010), that the formation of race categories is related to the emergence of the other-race effect. Specifically, the establishment of race categories may lead infants to restrict their estimation of extra-

personal appearance variation to within-category comparisons, leading to the emergence of a robust other-race effect. The maintenance of distinct racial norms (Armann et al., 2011; Short, Hatry, & Mondloch, 2011) may make these comparisons difficult to enact, for example, leading to training data that does not include data about cross-race appearance variability. That the model exhibited the other-race effect is less important than the potential insight that changing category structure during development may play an important computational role in how infants perform in the VPC scenario. This result raises an important question, however: where do categories come from? The restriction of cross-race comparisons in my previous simulations required an external teacher (Balas, 2012)—there is nothing built into the model to suggest that some difference images are different than others. What might lead an infant to determine that categories defined by race, sex, or age exist? I continue by demonstrating that the current model may be able to provide a useful answer to this question.

Here, I applied the model to another testing paradigm designed to determine the extent to which infants maintain representations of face or object categories (Quinn, Eimas & Tarr, 2001). Similar to the basic VPC paradigm described previously, infants in categorization tasks are familiarized to one set of stimuli, and their discrimination ability is measured via their looking time to test images that either match or do not match the familiarization images. In this paradigm, however, infants are familiarized with several unique images drawn from a common category (e.g. Asian faces) rather than a single stimulus. Next, they are tested by measuring their looking time to novel exemplars from the same category and novel exemplars from a different category (e.g. White faces). A bias for the test images belonging to the novel category suggests that the categories under consideration are represented by the infant visual system. A simple bias for new identities should lead to both types of test stimulus being highly salient and successful performance therefore should depend on the ability to recognize that there is a shared category label between the familiarization and same-category test images.

This paradigm has been successfully used to test the category structure of race in infant face perception (Anzures et al., 2010). Anzures et al. reported that Caucasian infants familiarized to images from one race category (either Caucasian or Asian faces) showed increased looking time to novel category test exemplars, with the additional caveat that the data also revealed an interesting asymmetry: Category familiarization with Caucasian faces led to less robust novelty preferences than Asian faces in 6-month-olds, though by 9 months, this asymmetry was not evident. Here, I attempt to model this behavior, which is strong evidence for the formation of discrete race categories in infancy. Does this behavior emerge naturally from the current model, or do we need to identify some other source of category information that ultimately drives the other-race effect in infancy?

Training Data

As in my previous simulations, I tested the model using White and Asian faces from the FERET database (Phillips et al., 1998). The images used here were cropped and aligned images of Caucasian and Asian faces that are part of the CSU Face Identification Evaluation System (Bolme et al., 2003), a computer-vision benchmark for face recognition algorithms. All stimuli were 150×130 pixels and depicted the face inside a cropped oval that included the jawline but excluded the hair. Face alignment is based on the position of the eyes, so other features are not perfectly matched across pictures. I excluded images of individuals that had visible piercings, abnormal facial pigmentation (e.g. large moles or other blemishes), yielding 600 White individuals and 130 Asian individuals. There were at least two pictures of each individual in the final stimulus set.

Training the model

I chose to simulate the performance of two groups of infants, one group with Asian face experience and another with White face experience. In each case, I estimated intra- and extra-personal variation using difference images obtained from samples of 16 and 64 individuals. Intra-personal differences were obtained by sampling one difference image from each individual, and extra-personal differences were obtained by sampling eight difference images for each individual in the database. We used principal components analysis to estimate the eigenvectors of each face space and embedded the training data from each class within a three-dimensional space. The embedded data points were then used to estimate the mean and variance of a normal distribution so that we could estimate the necessary likelihoods for new training images.

Testing the model

I tested the model by simulating individual trials of the categorization task described above, which requires that we select a set of images that will be treated as the stimuli an infant would be familiarized with as well as an additional set of images they would be shown during a test phase. The training data given to the model thus represents an approximation of an infant's visual experience prior to performing the hypothetical task, while the images we select to simulate a trial represent the stimuli that they would see during completion of the task. Training the model is thus distinct from selecting images to simulate the familiarization phase of the infant categorization task. We simulated categorization trials using a set of 8 familiarization images drawn from one race, and two test images from each race. None of these images were used as part of the training data that served as the basis for eigenface estimation and distribution fitting. Behaviorally, familiarization images are shown to each infant and the critical test is whether or not they prefer to look at novel images from a new category rather than novel images from the same category as the familiarization stimuli. To operationalize this within the context of the current model, I chose to compute the probability that each possible image difference arising from a test image and a familiarization image was an instance of extra-personal variation (p_{extra}). This yields 16 values of p_{extra} for the same-category and different-category exemplars. To summarize the novelty of the different-category test images relative to the same-category test images, I chose to express the preference for the novel category using the difference between the average value of p_{extra} for different-category comparisons and same-category comparisons. Formally, we evaluate the expression below to determine if an individual trial is correct or not.

$$\frac{1}{16} \sum_{i=1}^2 \sum_{j=1}^8 p(\Omega_E | \Delta_{ij}^{\text{different}}) > \frac{1}{16} \sum_{i=1}^2 \sum_{j=1}^8 p(\Omega_E | \Delta_{ij}^{\text{same}}) \quad (5)$$

I carried out 100 simulated trials for each of the following experimental conditions: (1) White or Asian training data, (2) 16 or 64 individuals used to generate training data, (3) White or Asian faces used as category familiarization. This yielded 8 unique simulations.

Results

Each set of simulations yields two measures that I use to describe the extent to which race categories are evident after learning intra- and extra-personal variation in the manner prescribed by the model. First, on each of the 100 trials I have simulated, we can determine whether or not the average value of p_{extra} is larger for the comparisons between the different-category exemplars and the familiarization stimuli than for the comparisons involving the same-category exemplars. This is a binomial variable that we anticipate will

exceed a 0.5 proportion if race categories are an emergent property of the model. Second, on each trial, we can also compute the magnitude of the difference between the two average values of p_{extra} . This is a continuous variable that we anticipate will exceed 0 if race categories are an emergent property of the model. In each case, the property that we hope to see reflected in the model's estimates is that new exemplars from a new category should look more different to the model than new exemplars from the old category.

Considering the proportion of correct trials in each simulation first, we find that, in general, the model assigns a larger average p_{extra} value to the different-category exemplars. In Table 1, I include the proportions obtained in each simulation. In nearly all cases, the ratio of "correct" trials significantly exceeds 0.5 (as assessed by a two-tailed binomial test). There are two exceptions to this: When 16 training individuals were used and White faces were used in the category familiarization trials, we find no preference for the test images from the novel category. In all other conditions, proportions of correct responses were significantly greater than chance.

Characterizing the data using a binomial test is analogous to using a forced-choice procedure for coding infant visual preference, and I complement this analysis of the model with an additional assessment of the magnitude of the difference between the average p_{extra} scores assigned to the two test conditions. In Figure 3, I display the average difference between the mean p_{extra} scores assigned to the different category and same category test images. To facilitate easy comparison relative to chance (a difference score of zero), I used the data from each run of 100 simulated trials to generate bootstrap estimates of the median difference and a 95% confidence interval for the median. Similar to the results described above for the analysis of proportion correct, the model assigns difference scores that are significantly greater than zero in all simulations with 64 individuals used for training. When only 16 individuals were used, I find that the model does not reliably assign significantly larger p_{extra} scores to different category exemplars, specifically failing to do so when the race of the training data is the same as the race of the category familiarization stimuli.

Overall, the data suggest that the model does behave in a manner that is consistent with infant perception of race categories. This suggests that the same model I have used to describe the onset of the other-race effect in infancy via category formation also provides an account of where those categories come from. Race categories, as assessed by this particular experimental paradigm, appear to emerge naturally from the model as a consequence of the estimates of intra- and extra-personal variation that are formed during the first year of life.

Discussion

The results of these simulations are interesting for several reasons. First, it is evident that performance in the categorization task is an emergent property of the model. Familiarization with faces of one race leads to a difference in p_{extra} that favors new individuals that also belong to a novel race category. This tendency is not robust for small amounts of training data, however – the results of the simulations indicate that the effect is not reliable when we used only 16 individuals to estimate intra- and extra-personal variation, but is reliable when we use 64 individuals. Further, we see some evidence of intriguing asymmetries in the model's results. It seems that performance (as assessed by forced-choice coding) in the category task is more robust when the model is familiarized with an Asian face category. This is similar to the result reported by Anzures et al. (2010), and extends their results by demonstrating that per the model, the dominant race in the infant's experience does not matter – familiarization with a Caucasian category leads to a smaller preference regardless of the dominant race in the training data. This is an interesting prediction for future empirical studies, suggesting that there is intrinsic asymmetry in the variability of facial

appearance across races. I note however, that the model produces a different pattern of results when we use the average difference in p_{extra} values as our metric of model performance. Given this, the strongest conclusion we can make from the current data is simple, but powerful – discrete race categories may emerge naturally from estimating intra- and extra-personal variation in a biased sample of faces.

There are a number of important challenges to address with regard to the computational basis of perceptual narrowing in the context of the other-race effect. Perhaps the most difficult to address with the current model is the challenge of accounting for data obtained with visual preference methods as opposed to VPC or familiarization paradigms. In typical infant visual preference tasks, experimenters code infant looking behavior in terms of the proportion of looking time allocated to one of two simultaneously presented patterns. Since the current model relies on difference images, the best approximation of these tasks we can achieve is to directly compare the two test images presented on a given trial and ask the model to provide an estimate of p_{extra} for the differences in appearance observed between those inputs. Unfortunately, while this would be adequate for modeling the extent to which different test pairs were discriminable from one another (e.g. predicting whether or not any visual preference should obtain), predicting which stimulus in a test pair will be preferred is beyond the scope of the current model. Infants exhibit well-defined preferences for own-race faces in early infancy (Kelly et al., 2005) that may be the first behavioral indices of the other-race effect, making this an important problem to solve if we wish to develop a comprehensive theory of the development of other-race face perception. Augmenting the model is not out of the question, but would require us to add several new features. A potentially easier target for further study is to use the current model to make predictions about eye movements to images of own- and other-race faces during infancy. Infants scan own- and other-race faces differently (Wheeler et al., 2011; Xiao et al., 2012), suggesting that they are perhaps learning different perceptual strategies for information pick-up that lead to more or less effective face recognition. The current model could be used to account for these biases in eye movements by defining areas of interest (AOIs) that match previous eye-tracking studies and examining how costly it is to add visual noise within different regions of the face. To the extent that there is systematic variability in accuracy as a function of AOI, this could be a useful way to define a saliency map that is based on the discriminability of intra- and extra-personal variability.

How does the model relate to visual learning?

Besides its utility as a computer vision model, this model of the other-race effect is also a useful foundation for understanding the computational basis of the other-race effect insofar as several aspects of its design are commensurate with recent results describing visual learning in the human visual system. In particular, I suggest that the overall architecture of the model, particularly its reliance on difference images, relates easily to several known properties of infant and adult visual learning. However, the relationship between real visual learning and the manner in which we have trained the model here also raises important issues for future work.

The current model relies heavily on supervised learning – without labeled instances of intra- and extra-personal variation, the model would have no means of building either of the dual eigenspaces necessary for classifying new patterns, nor would it be possible to estimate probability distributions that could be used to measure the likelihood that new inputs were instances of one class or the other. This may seem like a weakness of the model, since it is a reasonable question to ask what the learning signal is that makes supervised learning possible. To function properly, the model assumes that there is some means of accumulating instances of intra- and extra-personal variation and assigning them to the correct classes with

enough reliability that the learner is not faced with eigenspaces (and corresponding distributions) that are too similar. There are solutions to this problem that do not rely solely on visual input: If we believe that infants are capable of using information from other sensory modalities as a learning signal, then an individual's voice (Beauchemin et al., 2011) or any other external cue could be used as a learning signal to determine when two faces differ intra- or extra-personally. Indeed, external cues to identity are known to play a critical role in the efficacy of training regimens to extend the window of perceptual narrowing. The other-species effect, for example, can be postponed by providing infants with pictures of monkeys that have individual names accompanying them in a picture book (Scott & Monesson, 2009). Mere exposure to those same images with a common label accompanying them (e.g. "monkey") does not appear to improve performance (Pascalis et al., 2005). This finding generalizes to other-race faces (Heron-Delaney et al., 2011), suggesting the generality of individuation training in postponing the emergence of other-group face recognition effects. In terms of the model, these verbal cues may serve as a teaching signal that distinct images either represent instances of intra-personal variation (one "monkey" label) or extra-personal variation (unique labels for individuals), leading to the construction of a more or less effective representation of the distribution of appearance differences in the other-species category. The success of individuation training as opposed to category training (or mere exposure) in adults (Scott et al., 2008) and infants, suggests that explicit cues to identity are vital for face learning. Besides these non-visual teaching signals that could support learning per the current model's architecture, I also argue that recent results in adult visual learning suggest that there are sensible solutions to this problem that do not require an external teaching signal and make interesting predictions about how visual learning proceeds during perceptual narrowing.

Temporal contiguity is an important characteristic of our visual world that can serve as a teaching signal to support the acquisition of robust representations of appearance. Though most tests of the other-race effect use static images, the representations underlying performance in these tasks were almost certainly learned from moving stimuli – faces are rarely seen in the absence of motion, and dynamic facial appearance may provide a highly useful means of learning distributions of intra- and extra-personal variation. Temporal contiguity between distinct images is known to impact adult object learning (Wallis, 2002). If distinct images appear close together in time, this is a strong cue that those images belong to the same object and the same label should be applied to them. Neurons in primate inferotemporal cortex use temporal contiguity to generalize their responses over arbitrary patterns that occur close to one another in time (Miyashita, 1988), and the tendency to generalize object identity over short temporal windows can be used to "break" invariant recognition in adult observers (Cox et al., 2005). For example, Wallis & Bulthoff (2001) trained observers with a set of rotating faces that changed identity as the face turned from a profile view to a frontal view. The temporal contiguity between these images led individuals to learn a representation of the training faces in which the same label was attached to images of two different individuals, thus co-opting a typically useful means of learning invariant recognition so that an erroneous representation of appearance was learned. Cox et al. (2005) demonstrated that similar assumptions about temporal contiguity and object constancy could be used to break translation invariance. In their task, observers viewed objects in the periphery before making a saccade to bring the target object onto the fovea. Mid-saccade, object identity changed in some conditions, leading observers to learn a representation of object appearance in which two different images were bound to the same label. Both experiments demonstrate that the visual system uses temporal smoothness as a fundamental assumption that supports the acquisition of new object representations. With regard to the problem posed by the current model, I suggest that besides its role in developing the ability to generalize class information over distinct inputs, temporal contiguity may also be used to identify instances of intra-personal variability. Observers who are exposed to dynamic

objects do not just learn to generalize object labels across distinct images, but also exhibit increased sensitivity to different frames in a sequence (Balas & Sinha, 2008). Temporal contiguity thus serves a dual purpose: Observers can learn that the same label can be generalized across the images in a coherent sequence, and as a corollary they can learn that the observed changes in appearance comprising that sequence are incidences of intra-object variation. Thus, simply by observing an object (or a face) over an extended period of time, an observer can confidently accumulate examples of intra-personal variation so that half of our learning problem can be solved. What about the other half of the problem? How can observers also accumulate instances of extra-personal variability? There is little data that speaks directly to this problem, but I suggest that in natural environments, eye movements may provide a different teaching signal that strongly indicates that two different objects have been seen. Eye movements appear to “reset” perception (Paradiso et al., 2012) and a saccade to a different face may provide another teaching signal that the difference between the faces before and after the saccade are instances of extra-personal variation.

This proposal makes empirical predictions regarding the role of eye movements in visual learning that to my knowledge have not been tested. However, it also raises important questions about how the current model is trained and how this relates to the real learning scenario in infancy. If intra-personal variation is indeed learned by using temporal contiguity as a learning signal, then I suggest that it is highly likely that infants obtain far more samples of intra-personal variation than extra-personal variation. This is exactly the opposite of the current model’s experience, due to the lack of databases containing a large number of images from each individual. As described above, any database of images will have many more potential instances of extra-personal variation than instances of intra-personal variation, but varying the intra/extra ratio during training may be an important test of how the model maps onto real visual learning. In general, aligning the training data more closely with known results about real infant experience (Rennels & Davis, 2008) is an important direction for future work so that we can use the model to make clear quantitative predictions about behavior. Recent data from infant-mounted webcams (Sugden & Moulson, 2012) may be especially useful in this regard, offering quantitative estimates of how faces are tracked by infants.

Plasticity and other-race face recognition

The current model offers one account of how face recognition changes during infancy as a function of exposure to a predominance of faces belonging to one race. Besides considering the development of the other-race effect during this time period, the model may also yield insights into how own- and other-race face recognition change if experience changes significantly over the life-span. The other-race effect is known to be malleable in children (Sangrigoli et al., 2005; Bar-Haim et al., 2006, de Heering et al. 2010; Anzures et al., 2012) and adults (Degutis et al., 2011), and recent results indicate that chimpanzees’ face recognition abilities for human and chimpanzee faces undergo an extended developmental trajectory (Dahl et al., 2013) – at early ages chimpanzee faces are recognized more effectively than human faces (assuming that this is the dominant category in the environment, (Sugita, 2008)), but in later life (and with extended experience), human faces are ultimately favored. I do not know of any data from human observers describing the time-course of changes in the other-race effect over long periods of time, but the data from chimpanzees indicates that recognition accuracy obeys a fairly simple function. Is this also an emergent property of learning intra- and extra-personal variation in the manner described here? The current model makes it possible to at least crudely quantify experience by manipulating the amount and composition of training data, meaning we may examine how it responds to both acute changes in the visual environment and gradual changes in the diversity of experience available to the observer.

Conclusion

I have demonstrated how Bayesian classification of intra- and extra-personal face variation provides a useful model of the development of the other-race effect. The model relates easily to established paradigms for testing infant observers, uses real images as inputs, and facilitates the manipulation of visual experience in meaningful ways. Extending previous results obtained with the model, I have also shown how the model may account for the emergence of race categories in infancy, which I argue is a critical precursor to the development of the other-race effect. This model, while obviously incomplete in many ways, nonetheless offers a promising framework for studying the computational basis of visual learning in multiple domains. Both the generality of the algorithm and its modest success at identifying critical mechanisms that underlie the other-race effect make it an attractive foundation for future simulations.

Acknowledgments

I would like to acknowledge support from COBRE Grant P20 RR020151 from the National Institute for General Medical Studies (NIGMS) and NSF EPSCoR Grant #EPS-0814442. I would also like to thank Alyson Saville & Carol Huynh for helpful comments on the manuscript.

References

1. Anzures G, Quinn PC, Pascalis O, Slater A, Lee K. Categorization, categorical perception, and asymmetry in infants' perception of face race. *Developmental Science*. 2010; 13:553–564. [PubMed: 20590720]
2. Anzures G, Wheeler A, Quinn PC, Pascalis O, Slater AM, Heron-Delaney M, Tanaka J, Lee K. Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology*. 2012; 101(2):124–136. [PubMed: 18639888]
1. Armann R, Jeffery L, Calder AJ, Rhodes G. Race-specific norms for coding face identity and a functional role for norms. *Journal of Vision*. 2011; 11(13):9, 1–14. [PubMed: 22072729]
2. Balas B. Using innate visual biases to guide face learning in natural scenes: A computational investigation. *Developmental Science*. 2010; 13:469–478. [PubMed: 20443967]
3. Balas B. Bayesian face recognition and perceptual narrowing. *Developmental Science*. 2012; 15(4): 579–588. [PubMed: 22709406]
4. Balas BJ, Sinha P. Observed object motion induces increased generalization and sensitivity. *Perception*. 2008; 37:1160–1174. [PubMed: 18853553]
3. Bar-Haim Y, Ziv T, Lamy D, Hodes RM. Nature and nurture in ownrace face processing. *Psychological Science*. 2006; 17:159–163. [PubMed: 16466424]
5. Beauchemin M, Gonzalez-Frankenberger B, Tremblay J, Vannasing P, Martinez-Mones E, Belin P, Beland R, Francoeur D, Carceller A-M, Wallois F, Lassonde M. Mother and Stranger: An Electrophysiological Study of Voice Processing in Newborns. *Cerebral Cortex*. 2011; 21:1705–1711. [PubMed: 21149849]
4. Bolme, D.; Beveridge, R.; Teixeira, M.; Draper, B. The CSU Face Identification Evaluation System: Its Purpose, Features and Structure. *International Conference on Vision Systems*; Graz, Austria. 2003. p. 304-311.
6. Caldara R, Abdi H. Simulating the other-race effect with autoassociative neural networks: Further evidence in favor of the face-space model. *Perception*. 2006; 35:659–670. [PubMed: 16836056]
7. Cox DD, Meier P, Oertelt N, DiCarlo JJ. “Breaking” position-invariant object recognition. *Nature Neuroscience*. 2005; 8:1145–1147.
8. Dahl CD, Rasch MJ, Tomonaga M, Adachi I. Developmental processes in face perception. *Scientific Reports*. 2013; 3:1044. [PubMed: 23304435]
9. de Heering A, de Liedekerke C, Deboni M, Rossion B. The role of experience during childhood in shaping the other-race effect. *Developmental Science*. 2010; 13:181–187. [PubMed: 20121874]

10. Degutis J, DeNicola C, Zink T, McGlinchey R, Milberg W. Training with own-race faces can improve processing of other-race faces: evidence from developmental prosopagnosia. *Neuropsychologia*. 2011; 49:2505–2513. [PubMed: 21570991]
11. Fantz RL. Visual experience in infants: decreased attention to familiar patterns relative to novel ones. *Science*. 1964; 146:668–670. [PubMed: 14191712]
12. Freedland RL, Dannemiller JL. Nonlinear pattern vision processes in early infancy. *Infant Behavior and Development*. 1996; 19:21–32.
13. Furl N, Phillips PJ, O'Toole AJ. Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*. 2002; 26:797–815.
14. Ge L, Zhang H, Wang Z, Pascalis O, Quinn PC, Kelly D, Slater A, Tian J, Lee K. Two faces of the other-race effect: Recognition and categorization of Caucasian and Chinese faces. *Perception*. 2009; 38:1199–1210. [PubMed: 19817152]
15. Goren CC, Sarty M, Wu PY. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*. 1975; 56:544–549. [PubMed: 1165958]
16. Hancock PJB, Burton AM, Bruce V. Face processing: human and perception and principal components analysis. *Memory and Cognition*. 1996; 24:21–40. [PubMed: 8822156]
17. Haque A, Cottrell GW. Modeling the other-race advantage with PCA. *Cog Sci*. 2004:899–904.
18. Heron-Delaney M, Anzures G, Herbert JS, Quinn PC, Slater AM, et al. Perceptual Training Prevents the Emergence of the Other Race Effect during Infancy. *PLoS ONE*. 2011; 6(5):e19858. [PubMed: 21625638]
19. Humphreys K, Johnson M. The development of face-space in infancy. *Visual Cognition*. 2007; 15(5):578–598.
20. Kellman PJ, Arterberry ME. Infant Visual Perception. *Handbook of Child Psychology*. 2007
21. Kelly DJ, Liu S, Lee K, Quinn PC, Pascalis O, Slater AM. Development of the other-race effect in infancy: Evidence towards universality? *Journal of Experimental Child Psychology*. 2009; 104:105–114. [PubMed: 19269649]
22. Kelly DJ, Quinn PC, Slater AM, Lee K, Ge L, Pascalis O. The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*. 2007; 18:1084–1089. [PubMed: 18031416]
23. Kelly DJ, Quinn PC, Slater Am, Lee K, Gibson A, Smith M, Ge L, Pascalis O. Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*. 2005; 8:F31–F36. [PubMed: 16246233]
24. Levin DT. Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1996; 2:1364–1382.
25. Miyashita Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*. 1988; 335:68–70. [PubMed: 3412461]
26. Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. *Pattern Recognition*. 2000; 33:1771–1782.
27. Nelson CA. The development and neural bases of face recognition. *Infant and Child Development*. 2001; 10:3–18.
28. O'Toole AJ, Deffenbacher KA, Abdi H, Bartlett JC. Simulation of 'other-race effect' as a problem in perceptual learning. *Connection Science*. 1991; 3(2)
29. O'Toole A, Abdi H, Deffenbacher KA, Valentin D. Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America, A*. 1993; 10:405–411.
30. O'Toole AJ, Deffenbacher KA, Valentin D, Abdi H. Structural Aspects of Face Recognition and the Other-race Effect. *Memory & Cognition*. 1994; 22:208–224. [PubMed: 8035697]
31. Pappas MH, Goldinger SD. A multi-dimensional scaling analysis of own- and cross-race face spaces. *Cognition*. 2010; 116:283–288. [PubMed: 20501337]
32. Paradiso MA, Meshi D, Pisarcik J, Levine S. Eye movements reset visual perception. *Journal of Vision*. 2012; 12(13):11, 1–14. [PubMed: 23241264]

33. Pascalis O, de Haan M, Nelson CA. Is face processing species-specific during the first year of life. *Science*. 2002; 296:1321–1323. [PubMed: 12016317]
34. Pascalis O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, Nelson CA. Plasticity of face processing in infancy. *PNAS*. 2005; 102:5297–5300. [PubMed: 15790676]
35. Phillips JP, Weschler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*. 1998; 16:295–306.
36. Quinn PC, Eimas PD, Tarr MJ. Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *Journal of Experimental Child Psychology*. 2001; 79:78–94. [PubMed: 11292312]
37. Rennels JL, Davis RE. Facial experience during the first year. *Infant Behavior and Development*. 2008; 31(4):665–678. [PubMed: 18554724]
38. Sangrigoli S, Pallier C, Argenti AM, Ventureyra VAG, de Schonen S. Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*. 2005; 16:440–444. [PubMed: 15943669]
39. Scott LS, Monesson A. The origin of biases in face perception. *Psychological Science*. 2009; 20:676–680. [PubMed: 19422630]
40. Scott LS, Pascalis O, Nelson CA. A Domain-General Theory of the Development of Perceptual Discrimination. *Current Directions in Psychological Science*. 2007; 16:197–201. [PubMed: 21132090]
41. Scott LS, Tanaka JT, Sheinberg DL, Curran T. The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*. 2008; 1210:204–215. [PubMed: 18417106]
42. Short LA, Hatry AJ, Mondloch CJ. The development of norm-based coding and race prototypes: An examination of 5- and 8-year-olds' face space. *Journal of Experimental Child Psychology*. 2011; 108:338–357. [PubMed: 20822777]
43. Simion F, Valenza E, Macchi Cassia V, Turati C, Umiltà C. Newborns' preference for up-down asymmetrical configurations. *Developmental Science*. 2002; 5:427–434.
44. Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*. 1987; 4:517–519.
45. Sugden, N.; Moulson, M. Here's looking at you! Quantification of quotidian face exposure in adults and infants; Poster presented at the Society for Research in Child Development Themed Meeting: Developmental Methodology; Tampa, FL. 2012 Feb.
46. Sugita Y. Face perception in monkeys reared with no exposure to faces. *PNAS*. 2008; 105:394–398. [PubMed: 18172214]
47. Turk M, Pentland A. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*. 1991; 3:71–86. [PubMed: 23964806]
48. Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*. 1991; 43:161–204. [PubMed: 1866456]
49. Valentine T, Endo M. Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology*. 1992; 44A:671–703. [PubMed: 1615169]
50. Wallis G. The role of object motion in forging long-term representations of objects. *Visual Cognition*. 2002; 9:233–247.
51. Wallis G, Bülthoff HH. Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*. 2001; 98:4800–4804.
52. Wheeler A, Anzures G, Quinn PC, Pascalis O, Omrin DS, Lee K. Caucasian infants scan own- and other-race faces differently. *PLoS One*. 2011; 6:e18621. [PubMed: 21533235]
53. Xiao WS, Xiao N, Quinn PC, Gizelle A, Lee K. Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development*. 2012

Figure 1.

Intra- and extrapersonal variation. Intrapersonal variation refers to differences in facial appearance that arise when identity is preserved, but age, expression, occlusion, or other properties change-comparing the central image to those on the left, for example.

Extrapersonal variation refers to differences in facial appearance that occur when identity differs, which may include comparisons between faces of different ages, sexes, races, or species (central image vs. rightmost images). A successful observer will effectively distinguish between these two kinds of variability, but may have limited training data with which to tackle the problem. To view this figure in colour, please see the online issue of the Journal.

Figure 2.

Examples of intra- and extrapersonal variation as implemented in the current model. Note that, compared to Figure 1, these images are all carefully aligned and normalized, which are necessary prerequisites for the model.

Figure 3.

A schematic of an individual VPC trial that an infant might see in a typical laboratory task. The images at top are the familiarization stimuli, whereas the test images depict a new individual and a new image of the familiar individual. In terms of the model, success on this task may be supported by the computation of the posterior probability that each of the two candidate difference images represent extrapersonal variation. To view this figure in colour, please see the online issue of the Journal.

Figure 4.

Simulation results for the categorization task described in the main text. When a small amount of training data is used, the model does not reliably exhibit categorization behaviour, but with a larger amount it succeeds.

Table 1

Proportion of simulated trials with a preference for the novel category stimulus. P-values indicate the significance of one-tailed binomial tests.

	Training Data = 16 Individuals	Training Data = 64 Individuals
White Experience, Asian Category	75%, $p < 0.001$	79%, $p < 0.001$
White Experience, White Category	38%, n.s.	60%, $p = .028$
Asian Experience, Asian Category	59%, $p = 0.045$	77%, $p < 0.001$
Asian Experience, White Category	41%, n.s.	61%, $p = 0.0175$