

MIT Open Access Articles

Potential downside of high initial visual acuity

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Vogelsang, Lukas et al., "Potential downside of high initial visual acuity." Proceedings of the National Academy of Sciences of the United States of America 115, 44 (October 2018): 11333-38 ©2018 Authors

As Published: <https://dx.doi.org/10.1073/PNAS.1800901115>

Publisher: Proceedings of the National Academy of Sciences

Persistent URL: <https://hdl.handle.net/1721.1/129594>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Potential downside of high initial visual acuity

Lukas Vogelsang^{a,b,1}, Sharon Gilad-Gutnick^{a,1}, Evan Ehrenberg^a, Albert Yonas^c, Sidney Diamond^a, Richard Held^{a,2}, and Pawan Sinha^{a,3}

^aDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bInstitute of Neuroinformatics, University of Zurich/ETH Zurich, 8057 Zurich, Switzerland; and ^cDepartment of Psychology, Arizona State University, Tempe, AZ 85281

Edited by Deborah Giaschi, University of British Columbia, and accepted by Editorial Board Member Marlene Behrmann September 18, 2018 (received for review January 18, 2018)

Children who are treated for congenital cataracts later exhibit impairments in configural face analysis. This has been explained in terms of a critical period for the acquisition of normal face processing. Here, we consider a more parsimonious account according to which deficits in configural analysis result from the abnormally high initial retinal acuity that children treated for cataracts experience, relative to typical newborns. According to this proposal, the initial period of low retinal acuity characteristic of normal visual development induces extended spatial processing in the cortex that is important for configural face judgments. As a computational test of this hypothesis, we examined the effects of training with high-resolution or blurred images, and staged combinations, on the receptive fields and performance of a convolutional neural network. The results show that commencing training with blurred images creates receptive fields that integrate information across larger image areas and leads to improved performance and better generalization across a range of resolutions. These findings offer an explanation for the observed face recognition impairments after late treatment of congenital blindness, suggest an adaptive function for the acuity trajectory in normal development, and provide a scheme for improving the performance of computational face recognition systems.

visual development | visual acuity | deep neural networks | spatial integration | sight restoration

This work was initiated by a serendipitous referral to our laboratory of a young boy, RK, with an unusual visual history. RK was born in China and placed in an orphanage soon thereafter. He had dense bilateral congenital cataracts, but lack of resources prevented him from receiving treatment until he was 4.5 y old. Two years after his surgery, an American family adopted and brought him to the United States, where he lives today. RK adapted well and his health and intellectual development progressed satisfactorily. However, his parents noticed that despite being otherwise visually proficient, he had difficulty recognizing people's faces, an impairment that was impeding his ability to socialize. They contacted us to request that we assess RK's vision. In the laboratory, we tested RK along multiple perceptual and neurological dimensions including acuity, contrast sensitivity, face/nonface discrimination, within-class object individuation, and face individuation. Consistent with his parents' observations, we found that RK performed normally on all tests except face identification, on which his responses were no better than chance.

A review of the literature reveals that RK is not unique in his performance profile; his recognition deficit echoes results that have previously been reported in studies with adolescent children who underwent treatment for congenital cataracts during the first year of infancy. They exhibit impaired face-discrimination performance even when tested several years after their treatments (1–3). Notably, the periods of deprivation they suffered were quite short, ranging between 2 and 6 mo.

The finding that even brief periods of visual deprivation can lead to face-identification deficits, while appearing to spare other visual skills, has been explained as the manifestation of a critical period for face learning (4, 5). According to this account, face recognition is an “experience-expectant” process, requiring exposure to faces early in development to enable appropriate

perceptual and cortical specialization necessary for face-identification abilities (1, 6–8). Children like RK, who pass this period without normal exposure to faces, are expected to exhibit compromised face recognition skills later in life. Recent evidence from nonhuman primates who have undergone controlled deprivation is consistent with these results (8). Monkeys reared without exposure to faces lacked face preference in their looking behavior, and did not exhibit face-specialized cortical domains, in contrast to nondeprived controls. These observations suggest that skills which appear early in the developmental timeline, such as face discrimination, may be particularly vulnerable to visual deprivation. It is worth pointing out, however, that the consequences of delayed treatment of congenital blindness are complex and do not all conform to a unitary template. Several high-level visual skills appear capable of being acquired even after prolonged periods of initial deprivation (9–12). Additional evidence of resilience is provided by studies of low-level vision. Here, findings suggest that the earliest appearing proficiencies of normal development, such as the ability to perceive visual flicker, are also the ones least susceptible to deprivation, a biological analog of the common corporate practice of “first-hired last-fired” (13, 14). Taken together, these studies point to a varied landscape of visual proficiencies following late treatment of congenital blindness, with some skills more susceptible to compromise by visual deprivation during early “sensitive” periods in development. Specifically, impairments of face processing following early visual deprivation may plausibly admit a sensitive

Significance

As newborns, we start with poor visual acuity, attributable to retinal and cortical immaturities. This has been considered to be a limitation of early visual processing. We propose that initially poor retinal acuity may, in fact, have adaptive value. It may help set up processing strategies and receptive fields in the cortex that facilitate spatial analysis over extended areas. Observations from children who started their visual journeys with abnormally high visual acuity, and computational simulations with deep neural networks trained on high-resolution or blurred images, corroborate this proposal. The results have implications for understanding normal visual development; identifying causal roots of, and interventions for, important visual processing impairments; and creating better training procedures for computational vision systems.

Author contributions: L.V., S.G.-G., E.E., A.Y., S.D., R.H., and P.S. designed research; L.V., S.G.-G., E.E., S.D., and P.S. performed research; L.V., S.G.-G., E.E., and P.S. analyzed data; and L.V., S.G.-G., A.Y., S.D., and P.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.G. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

¹L.V. and S.G.-G. contributed equally to this work.

²Deceased November 24, 2016.

³To whom correspondence should be addressed. Email: psinha@mit.edu.

Published online October 15, 2018.

period-based account. However, an explanation that is not dependent on the domain-specific action of visual deprivation would have the virtue of parsimony. Here, we consider a domain-general account and present results from computational tests of its predictions.

Rather than focusing exclusively on the absence of information in the period prior to treatment as a critical contributor to later face-processing impairments (the “experience-expectant” viewpoint), an alternative is to consider how visual experience in the period following treatment differs in children like RK relative to the typically developing newborn. An important source of such differences are maturational processes in the visual system that continue progressing even while the eye has an occlusive pathology like a cataract. A key dimension impacted by such processes is acuity (15–17). Typically developing newborns commence their visual experience with remarkably poor acuity, below 20/600 (18–20), which is well beyond the criterion for legal blindness. Much of this acuity impairment can be attributed to the immature neonatal retina (21, 22), with its reduced foveal photoreceptor packing density (23), as well as to immaturities in the visual cortex (24, 25). Over the initial months of development, these immaturities diminish steadily, leading to a stereotyped pattern of acuity improvement (26). In a child with congenital cataracts, the physiological mechanisms of neural maturation continue to proceed despite the lenticular opacity (27, 28). As a consequence of this maturational progression, once the cataracts are eventually removed, the initial retinal acuity of the child is significantly higher than that of a newborn. Indeed, this was the case with RK, whose visual experience started with an acuity of 20/40.

There are several notable aspects of the vision of children treated for congenital cataracts (29), including a lack of accommodation [given the fixed-focus aphakic correction comprising either intraocular lenses (IOLs) or external glasses], changes in ocular growth (30), and the development of deprivation-induced amblyopia (31, 32). The studies have spanned a spectrum of ages [with age at surgery ranging from less than 6 mo (33–35) to 8 y (36, 37)]. The data show that although outcomes are generally better the earlier the surgery is conducted, and long-term acuity does not reach normal levels, of particular relevance here is the finding that these children exhibit better than neonatal acuity rapidly after treatment.

Although high retinal acuity at the outset may appear to be advantageous for the visual system given the richer visual experience it offers, it may, in fact, compromise the development of important visual processes, specifically those involving extended spatial integration. Poor acuity, by definition, introduces blur that reduces effective image resolution. Small patches in blurred images do not contain enough structure to be informative; spatial integration is necessary to detect and discriminate patterns in such images (38). With high-resolution images, however, local processing suffices for discrimination tasks, rendering extended spatial integration superfluous (39).

To formalize this intuition, in a simple fragment-based image classification system (detailed in *Methods*), we examined fragment sizes needed in high-resolution versus blurred images to obtain similar levels of classification performance. As Fig. 1 shows, to achieve the same z-score (or comparable classification performance), a smaller integration area suffices for high-resolution images relative to blurred ones. The implication is that the early availability of high-resolution information may obviate the development of strategies for integrating over extended spatial extents. Such integration is a prerequisite for the analysis of configural relationships in images (40), which, in turn, are believed to be especially important for facial identification (41, 42). This line of reasoning, first suggested as a possible explanation for some of the observed “sleeper effects” following early visual deprivation (43), predicts that children with initially high acuity would be biased toward local processing. Indeed, this is consistent with data showing

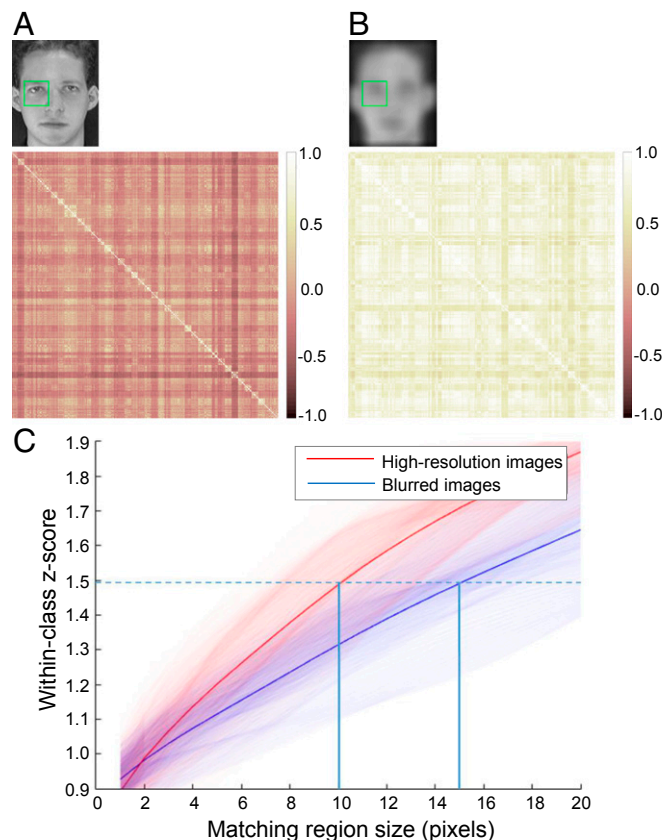


Fig. 1. (A and B) Confusion matrices derived from all pair-wise comparisons of 400 face images (10 exemplars per 40 individuals). The matrices are averaged across 10 matching regions distributed across the image (an example matching region is indicated by the boxes overlaid on the face images), when the images are processed at high resolution (A) or with blur (B). Lighter shades indicate higher match scores. The reduced distinctiveness of content in the blurred image patches leads to the confusion matrix in B being dominated by indiscriminately high match scores, relative to A. Note that the two confusion matrices are plotted using identical color scales. (C) From each confusion matrix, we can determine how different the “within-class” scores (for the 10 images of the person shown in the reference image) are relative to the scores of the other 390 images. The plots show summary results of 5,208 simulations using different matching region locations and sizes when the images are in high resolution (red curves) or blurred (blue curves). The bold curves represent the averages across all individual simulations (seen here as diffuse background curves). Within-class z-scores are plotted against region sizes. A higher z-score corresponds to better differentiation of within-class images relative to out-of-class images. As is evident from the plots, to achieve the same level of class discrimination, a smaller image fragment suffices for the high-resolution case compared with the blurred one. Face images courtesy of AT&T Laboratories Cambridge.

that children treated for congenital cataracts are able to perform face discrimination tasks where the patterns differ in local features but are impaired on a task requiring detection of configural changes (44), as well as on holistic face processing (45). In summary, the abnormally high initial retinal acuity experienced by children treated for congenital blindness later in life, compared with typical newborns, may result in compromised spatial integration. We refer to this as the high initial acuity (HIA) hypothesis.

To test the HIA hypothesis’ predictions about how experience with different acuity levels can affect the development of spatial integration fields and overall face-recognition performance, we trained different instances of a deep convolutional neural network (CNN) (46) on a large database of face images (47) while manipulating the blur of training and test data. It is worth

pointing out the analogies between the computational and biological entities implicit in this approach. First, image acuity corresponds to the resolution of the visual input to primary visual cortex, with limitations deriving primarily from retinal factors. This is in keeping with empirical data suggesting that retinal immaturities are significant determinants of observed visual sensitivity (48). In the CNN simulations, the resolution change is imposed at the input layer (akin to the output of the retina), and we examine how the response properties of units in the convolutional layers, which are posited to roughly correspond to primary visual cortex (49), change as a function of the input resolution. Pliability of cortical units is based on results from studies showing that visual deprivation can extend plastic periods (50, 51). The implication of these results is that with the onset of sight, some units in the visual cortex can develop differently depending on input properties, such as resolution. With this background, we examined how receptive fields (RFs) in the first convolutional layer and performance levels of a CNN change when image blur is varied. These tests yielded several notable results.

First, CNNs trained with image sets of different blur levels exhibited markedly different RFs in the first convolutional layer. As we had hypothesized, introduction of blur led to a progressive increase in the spatial extents of the RFs (Fig. 2*A* and *B*) ($P < 0.01$; two-tailed t -tests comparing RF sizes for each pair of consecutive blur levels). Second, as is clear from Fig. 2*C*, the presence of high spatial frequencies during training drives the system to behave in a biased way in which making use of those higher spatial frequencies is prioritized, while low spatial-

frequency information (necessary for effective generalization to blurred images) is mostly disregarded. The third result, also evident in Fig. 2*C*, is that none of these training regimens yields broad generalization performance. When testing the network with new images spanning a range of blurs, recognition accuracy peaks at the blur level matching the one used during training. This has an interesting implication: The front end of a network (comprising convolutional layers) trained with blurred images does not impose mandatory low-pass filtering on the inputs. If such filtering were to be employed, then high-resolution images, which subsume low spatial-frequency information, would have yielded the same performance as blurred inputs. Instead, high spatial-frequency information does progress through the network and influences eventual classification. That this influence is a detrimental one suggests that besides being able to utilize low-spatial frequency information, the network also needs to learn the appropriate weights for the high spatial-frequency components of the image data. Taken together, these results bear out our expectation of image resolution as a modulator of the extent of spatial integration by the initial RFs. They also point to the insufficiency of blurred images on their own for achieving broad generalization across resolution levels. More pragmatically, they reveal a limitation of CNN-based image recognition systems that are often trained on only high-resolution inputs (46, 52).

Motivated by the fact that visual experience in the normally developing brain involves a temporal progression from low to high acuity, we trained additional instances of the CNN using a staged regimen that commenced with blurred images, which were then followed by high-resolution ones. To examine ordering effects comprehensively, we trained four different instances of the CNN, one on each of the following regimens: “blurred-to-high-resolution” (250 epochs of training with blurred images followed by 250 epochs of training with high-resolution images); “high-resolution-to-blurred” (250 epochs of training with high-resolution images followed by 250 epochs of training with blurred images); “blurred-to-blurred” (500 epochs with exclusively blurred training images); and “high-resolution-to-high-resolution” (500 epochs with exclusively high-resolution training images).

These simulations yielded three primary results, summarized in Fig. 3. First, the introduction of blurred images in the training phase consistently induced the system to increase the sizes of its RFs, irrespective of when in the training regimen the blurred training was introduced (Fig. 3*A*). Interestingly, having high-resolution images follow the blurred ones did not shrink the size of the RFs. In other words, there is a notable asymmetry in terms of the effects of ordering: Starting with high-resolution images and then later introducing blurred ones leads to a significant increase in RF sizes, but the converse (blurred followed by high-resolution images) does not cause the network to reduce the sizes of the already established large RFs.

Second, even when the initial phase of training on blurred images is a small proportion of the full training regimen (in our simulations, 20%), it is sufficient to induce the creation of large RFs that are stable over the subsequent training epochs with high-resolution images. Increasing the amount of initial blurred training beyond this proportion does not cause a further expansion of the RFs (Fig. 3*B*). Drawing a parallel to human development, this observation suggests that although acuity improves rapidly in infancy (26), even the limited amount of time that a child experiences very poor acuity may well be sufficient for the instantiation and consolidation of RFs with large spatial extents.

Third, the order in which images of different resolutions are presented affects subsequent classification performance (Fig. 3*C*). Blurred-to-high-resolution training produces by far the most superior and generalized classification performance across all resolution levels, as indicated by it having the largest area under the curve (AUC) and lowest absolute slope (Fig. 3*D*). In contrast, when the network is trained on high-resolution to blurred, it yields poor performance when tested on high-resolution images, although, in

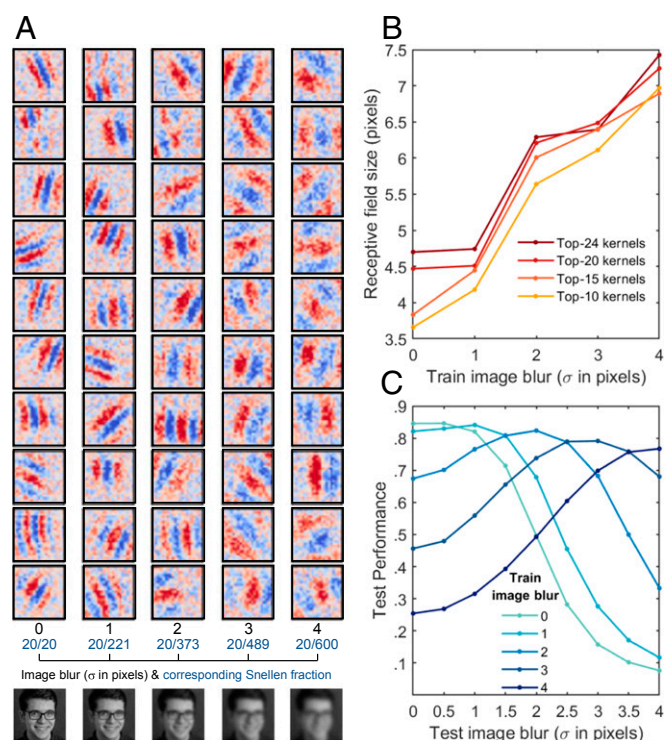


Fig. 2. Results of uniform training on CNNs. (*A*) Top-10 RFs in the first layer of CNNs trained on images blurred with a Gaussian filter with $\sigma = 0, 1, 2, 3, 4$, and corresponding acuities. The stronger the blur, the bigger and smoother the Gabor patches of the depicted RFs appear to be. (*B*) RF size increases with increasing blur. (*C*) Performance curves for training and testing CNN instances on different levels of blur. Performance is tied to the blur level of the training set, peaking at the test level that matches training level, although training on blurred images results in better generalization across other resolutions than training on high-resolution images. Face images courtesy of AT&T Laboratories Cambridge.

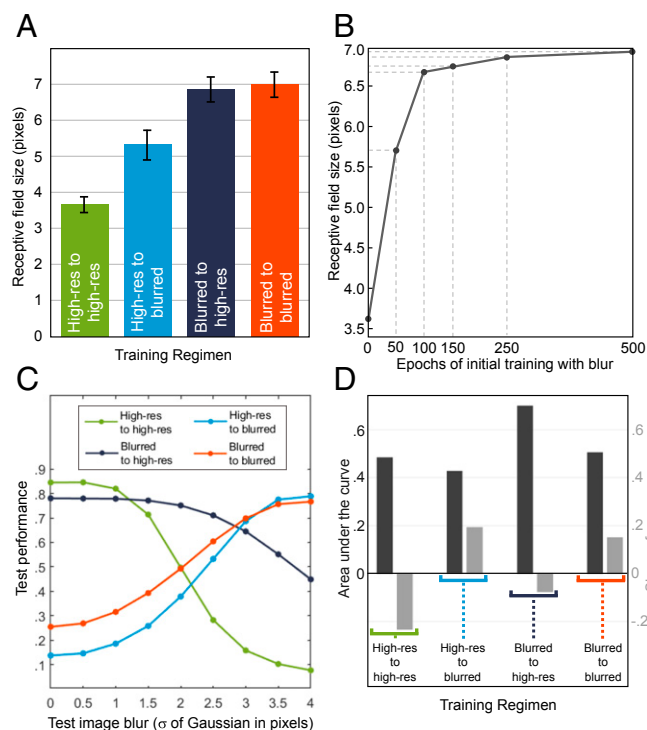


Fig. 3. Results of nonuniform training on CNNs. (A) Impact of nonuniform training paradigms on RF sizes. (B) When training paradigm begins with blur, even 20% blurred images are sufficient for the system to produce large RFs. (C) Impact of different training paradigms on performance level obtained with test images subjected to different levels of blur (here quantified as the σ , in pixels, of the convolving Gaussian): Blurred-to-high-resolution shows the best performance and generalization. (D) Two metrics of performance of the four different nonuniform training regimens. Black bars, area under the performance curve, normalized to 1.0 for perfect performance across all resolutions. Gray bars, slope of performance curve (perfect generalization would correspond to a slope of 0.0).

aggregate, it has been trained with precisely the same set of images as in the blurred-to-high-resolution condition. Thus, the superior performance of blurred-to-high-resolution training is a direct consequence of the initial period of blurred training, akin to the typically developing acuity trajectory of the human visual system.

A potential explanation for the pronounced ordering effects we observe is that the initial learning phase with only high-resolution images has a detrimental effect on generalization across resolution (Fig. 2C), causing a large training error in subsequent blurred learning trials, and finally a strong weight adaptation, even within the first convolutional layer. Thus, high spatial frequency features end up being replaced by those effective for classifying blurred images. This is not the case when the resolution progression is from low to high. Initial training on blurred images produces RFs with large spatial extents and superior generalization across resolutions; this diminishes the need for radical weight adaptation during follow-up training with high-resolution images. Corroborating this point, the blurred-to-blurred and blurred-to-high-resolution training regimens lead to similar RF sizes in the first convolutional layer.

Taken together, results from simulations with deep convolutional neural nets lend support to the proposal that in the developmental progression of visual acuity, initial experience with blurred imagery may help set up RFs capable of encoding image structure over extended spatial extents. This helps reduce reliance on local features and improves generalization performance across a range of image resolutions. The HIA account, while not ruling out the critical period-based explanation, can

parsimoniously account for the puzzling observations reported in the literature about early visual deprivation resulting in specific face-processing impairments.

Complementing data from the computational tests described here, the HIA hypothesis finds tentative support in the experimental literature. First, interventions like lid suture, which are akin to strong low-pass filtering, have been reported to result in enlarged RFs in the visual cortex (53). Second, recent studies of population RFs in human visual cortex (54) have reported an association between RF sizes and performance on face tasks, with smaller fields corresponding to compromised configural analysis.

The parsimony of the HIA hypothesis predicts that the consequences of sight recovery after early deprivation are not face-specific but should also apply to other tasks that rely on extended spatial processing. Empirical data support this prediction. Global motion processing is found to be compromised in this population even years after treatment (55). By contrast, patients who developed cataracts after one year show normal levels of global motion thresholds when tested several years after surgery. Individuals who undergo late treatment of congenital blindness have also been shown to be impaired at illusory contour perception and global shape completion but not on local shape discrimination (56). One study (57) found a greater deficit in a feature-spacing change detection task for faces relative to houses, but they attribute this impairment to compromised experience-dependent perceptual narrowing for faces that typically developing individuals exhibit. In a follow-up study (58), they highlight the idea that there may well be a general, rather than face-specific, mechanism used for spacing judgment tasks.

The results here suggest that the initially poor acuity in the normal developmental progression may be a feature of the system, rather than a limitation. Looking beyond acuity, it remains to be seen whether developmental progressions along other dimensions of visual function, such as field of view (59), color saturation (60), windows of simultaneity (61), and attentional control (62), might also confer adaptive benefits for instantiating robust processing strategies. In addition, there is abundant literature on the temporal dynamics of spatial-frequency processing, simulating HIA in computational models or in various behavioral, physiological, and imaging studies on normal adults. This body of research establishes the prevalence of coarse-to-fine sequencing at all levels of cortical processing of sensory information and across different modalities of sensation. For example, a study of temporal integration of rapidly sequenced spatially filtered visual images (63) found that success of the integration process in a normal adult population depended on a coarse-to-fine order of frame presentation. An fMRI study of temporal dynamics of face processing in higher-level visual cortex (64) varied exposure duration and spatial frequency content of filtered images. They demonstrated a low-frequency response to short duration exposures in a number of face-responsive regions of adult subjects, more robust and decaying more slowly with increasing exposure durations in bilateral fusiform face area (FFA), rFFA > lFFA. Other work (65, 66) has examined the benefits of adopting a “coarse to fine” matching strategy especially in the context of binocular stereopsis. Simulations suggest that such a progression might be beneficial for fine-tuning disparity sensitivity in binocular RFs (67). The researchers compared the results of simulated learning in networks, trained with four different sequences of filtered input, to detect binocular disparities in pairs of visual images. They found what they referred to as “developmental” sequences, namely those initial training sequences that employed exclusively either low or high frequency images, consistently outperformed the others in which the spectral content in the first stage of training was either random or identical to the following one.

The adverse effects of HIA may also have relevance beyond the domain of vision. For instance, a fetus’ auditory experience in the womb comprises low-pass filtered versions of voices and other

sounds in the external world (68). This kind of input may induce the development of extended temporal integration mechanisms so as to be able to detect envelope modulation of the auditory inputs. Significantly premature birth limits this low-frequency exposure and immerses the baby in an auditory environment teeming with high temporal frequencies. In a manner analogous to what we have proposed for high initial visual acuity, this abnormal auditory experience may lead to compromised hearing skills. Interestingly, premature birth is indeed found to be associated with higher order auditory impairments (69).

From the perspective of computer vision, our results point to an improved training strategy for enhancing the generalization performance of current CNN-based recognition systems and, more generally, illustrate the potential benefits of incorporating aspects of human developmental trajectories in designing training routines for machine-based systems.

The HIA hypothesis of potential benefits of exposure to reduced resolution imagery early in the developmental timeline raises interesting questions with basic and applied implications. A particularly far-reaching one regarding clinical practice relates to the issue of refractive correction following cataract surgery: When treating infants with cataracts, might it be better to leave their eyes without aphakic correction? We believe that a few considerations argue against changing the standard approach of providing aphakic correction, either via intraocular lenses or external glasses. First, uncorrected aphakia significantly degrades image resolution beyond the acuity limitations imposed by retinal immaturities in infants' eyes. This is especially true of newborn eyes given that their mean lens power is significantly greater than that of adults (34.4 D vs. 18.8 D) (70). The excessively degraded images in the absence of lenticular correction may prove inadequate for helping to define RF structures. Interestingly, from this perspective, the level of image blur experienced by a typical newborn may be seen as occupying a "Goldilocks" spot: not too high, to prevent shrinkage of RFs, and not too low to contain enough image structure that can guide the development of RF morphology. Second, our simulations indicate that the period of low acuity needed to support extended spatial integration is quite short and needs to be followed by a period of improved acuity. Just extending the duration of poor vision by failing to provide aphakic correction would not be sufficient for obtaining classification performance gains. Also, leaving an eye without aphakic correction might set in motion mechanisms for changing eyeball shape that might permanently compromise visual quality (71). For these reasons, leaving an eye without aphakic correction might not be a recommended course of action following cataract surgery. However, it may be worth considering the option of initially undercorrecting the refractive errors and progressively moving toward full correction. In the context of ongoing efforts to provide sight-initiating surgeries for congenitally blind children (72), our results suggest that patients with high postoperative acuity (31) might benefit from a temporary period of optical blur via refractive undercorrection to induce the visual system to develop appropriate spatial integration strategies.

Broadening our scope to include children born without any ocular pathologies, the HIA hypothesis induces us to consider whether newborns who start with better-than-typical acuity might be at risk for developing impairments in configural analysis and, specifically, face recognition. A longitudinal study of many children, whose acuity can be characterized at birth and whose configural processing can be assessed a few years later, can help detect such a link.

Methods

Simulation 1 (Results Presented in Fig. 1). We used 400 112×92 pixel grayscale face images taken from the AT&T ORL Database (73), comprising 10 instances of 40 individuals, allowing for matching across appearance transformations. All images were duplicated and subjected to a nine-pixel blur kernel such that the final set of 800 images consisted of two copies of each image: one in high-resolution and one blurred. For each sample image that was used to query the rest of the

database, we matched image patches from 2,604 locations, with 20 box sizes per location. For each image and face patch, we calculated the distance (L1 norm) between the given face patch, F_i , and all other corresponding face patches in each of the remaining images. Ideally, this returns distances to other instances of the same ID ("within ID" values) that are smaller than the distances to all other faces ("across ID" values). The z-score allows to compare these within-ID distances with the across-ID distances; the larger the gap between the average distances, the larger is the corresponding z-score for a given patch. Repeated across all images, this generates a distance matrix from which one can extract the average within-class z-score across all identities, for a given face patch. This is averaged across all patch locations to determine aggregate z-scores. For each image, at every location that a face patch is sampled from, we parametrically change the patch size to calculate the z-score as a function of query box size. By comparing these functions across high-resolution vs. blurred images, we characterize how different patch sizes contribute to achieving the same z-scores depending on image resolution.

Simulation 2.

Data for training and testing the CNN. A total of 50,429 face-cropped color images (100×100 pixels) belonging to 388 different face identities (average = 130 ± 20.4 images per class) were used for training and testing our CNN. The data were provided by the Facescrub image database (47), and all identities with less than 100 available images were rejected. Images were converted from color to grayscale, zero-centered to the overall mean luminance, and scaled by the overall SD. During training, images were randomly left-right flipped and rotated to up to 25° . For blurring images, we applied a Gaussian filter with $\sigma = 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$. The original images ($\sigma = 0$) are termed high-resolution. At a viewing distance of 60 cm, a head subtends $\sim 15^\circ$ of visual angle. An individual with 20/20 acuity would be able to resolve 450 cycles in this extent, while a person with 20/600 acuity could resolve 15 cycles. With this basic constraint and given a face image width of w pixels (the specific value of w is dictated by the requirements of the input layer of the CNN), we computed a Gaussian filter that removed spatial frequencies higher than $1/(w/15)$. Convolution of the image with this filter yielded an image approximating the effects of 20/600 acuity. The same procedure was applied for other acuity values. The data were split into 45,385 training and 5,044 test images aiming at a 90/10% training/test data split while keeping a constant number of test instances per class.

Network architecture and parameters. We utilized the AlexNet architecture (46), adjusted the network's input size to 100×100 pixels, and changed the shape of the 96 kernels in the first layer from 11×11 to 22×22 pixels to allow more reliable RF analyses. The net contains five convolutional layers (ReLU activation function), followed by three fully connected layers (tanh activation function; last layer equipped with softmax, crossentropy loss function and momentum optimizer). In between fully connected layers, a dropout of 50% is built into the network. The CNN was implemented in the TensorFlow deep learning library TFLearn and trained on a single GPU using a fixed learning rate of 0.001, minibatch learning with batch sizes of 128 and 500 epochs. The CNN instances were trained on images in high-resolution and several different blur levels. In addition, we utilized a high-resolution-to-blurred configuration by training a CNN instance first on high-resolution and later on blurred images (for 250 epochs each). Training an instance in reversed order provided us with a blurred-to-high-resolution configuration.

Weight analysis. For analyzing the RFs of the CNN, we investigated the shape of the Gabor wavelets in the kernels of the first convolutional layer. To quantify their geometric properties, we considered the two most dominant neighboring ellipses, one containing positive and the other containing negative values. We defined the separation between these two ellipses as the critical metric capturing the RF size. We applied thresholding, erosion, and dilation to extract individual ellipses and separated this process for positive and negative weights. For both, we only kept the detected ellipses with largest area. We then determined the two points of largest Euclidean distance in each region. To measure RF size, we used the distance between the middle point of the two points of the smaller ellipses and the line spanned by the two extreme points of the bigger ellipses. The procedure we use for fitting ellipses to the CNN "RFs" to characterize their structure and size is analogous to approaches adopted in neurophysiology wherein parametric forms, such as Gaussian or Gabor functions, are fit to empirically recorded RF maps (74, 75). We decided to fit each subfield with an ellipse, rather than one functional form to the overall RF, so as not to impose any preconceptions on the overall CNN RF organization.

Since some but not all of the 96 kernels in the first layer depicted a reliable Gabor pattern, we took into account the 10, 15, 20, and 24 kernels showing the most dominant structure (assessed using SD). In all configurations, the color translation has been scaled by a fixed value representing the inner 99% of the histogram of all pixel values in the first layer of the CNN instance trained on high-resolution.

ACKNOWLEDGMENTS. We thank Drs. Shlomit Ben-Ami, Rachel Robbins, Bas Rokers, Jitendra Sharma, and Galit Yovel and specially acknowledge the contributions of late Dr. R.H. to this work, and also, more broadly, the

fields of visual perception and development. This work was supported by the Nick Simons Foundation and National Eye Institute Grant EYR01020517 (to P.S.).

- Geldart S, Mondloch CJ, Maurer D, de Schonen S, Brent H (2002) The effects of early visual deprivation on the development of face processing. *Dev Sci* 5:490–501.
- Putzar L, Hötting K, Röder B (2010) Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restor Neurol Neurosci* 28:251–257.
- de Heering A, Maurer D (2014) Face memory deficits in patients deprived of early visual input by bilateral congenital cataracts. *Dev Psychobiol* 56:96–108.
- Rivolta D (2013) Cognitive and neural aspects of face processing. *Prosopagnosia* (Springer, Berlin), pp 19–40.
- Röder B, Ley P, Shenoy BH, Kekunnaya R, Bottari D (2013) Sensitive periods for the functional specialization of the neural system for human face processing. *Proc Natl Acad Sci USA* 110:16760–16765.
- de Schonen S, Mathivet H (1989) First come, first served: A scenario about the development of hemispheric specialization in face recognition during early infancy. *Eur Bull Cogn Psychol* 9:3–44.
- Nelson CA (2001) The development and neural bases of face recognition. *Infant Child Dev* 10:3–18.
- Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS (2017) Seeing faces is necessary for face-domain formation. *Nat Neurosci* 20:1404–1412.
- Ostrovsky Y, Andalman A, Sinha P (2006) Vision following extended congenital blindness. *Psychol Sci* 17:1009–1014.
- Ostrovsky Y, Meyers E, Ganesh S, Mathur U, Sinha P (2009) Visual parsing after recovery from blindness. *Psychol Sci* 20:1484–1491.
- Held R, et al. (2011) The newly sighted fail to match seen with felt. *Nat Neurosci* 14:551–553.
- Gandhi TK, Ganesh S, Sinha P (2014) Improvement in spatial imagery following sight onset late in childhood. *Psychol Sci* 25:693–701.
- Levi DM, Carkeet AD (1993) Amblyopia: A consequence of abnormal visual development. *Early Visual Development: Normal and Abnormal*, ed Simons K (Oxford Univ Press, New York), pp 391–408.
- Lewis TL, Maurer D (2009) Effects of early pattern deprivation on visual development. *Optom Vis Sci* 86:640–646.
- Huttenlocher PR, de Courten C, Garey LJ, Van der Loos H (1982) Synaptogenesis in human visual cortex—Evidence for synapse elimination during normal development. *Neurosci Lett* 33:247–252.
- Banks MS, Crowell JA (1993) Front-end limitations to infant spatial vision: Examination of two analyses. *Early Visual Development: Normal and Abnormal*, ed Simons K (Oxford Univ Press, New York), pp 91–116.
- Wilson HR (1993) Theories of Infant Visual Development. *Early visual development: Normal and abnormal*, ed Simons K (Oxford Univ Press, New York), pp 560–569.
- Dobson V, Teller DY (1978) Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Res* 18:1469–1483.
- Sokol S (1978) Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Res* 18:33–39.
- Courage ML, Adams RJ (1990) Visual acuity assessment from birth to three years using the acuity card procedure: Cross-sectional and longitudinal samples. *Optom Vis Sci* 67:713–718.
- Banks MS, Bennett PJ (1988) Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. *J Opt Soc Am A* 5:2059–2079.
- Candy TR, Banks MS (1999) Use of an early nonlinearity to measure optical and receptor resolution in the human infant. *Vision Res* 39:3386–3398.
- Yuodelis C, Hendrickson A (1986) A qualitative and quantitative analysis of the human fovea during development. *Vision Res* 26:847–855.
- Jacobs DS, Blakemore C (1988) Factors limiting the postnatal development of visual acuity in the monkey. *Vision Res* 28:947–958.
- Kiorpes L, Movshon JA (2004) Neural limitations on visual development in primates. *The Visual Neurosciences*, eds Chalupa LM, Werner JS (MIT Press, Cambridge, MA), pp 158–173.
- Daw N (2014) *Visual Development* (Springer, New York), 3rd Ed.
- Boas JAR, Ramsey RL, Riesen AH, Walker JP (1969) Absence of change in some measures of cortical morphology in dark-reared adult rats. *Psychon Sci* 15:251–252.
- Hendrickson A, Boothe R (1976) Morphology of the retina and dorsal lateral geniculate nucleus in dark-reared monkeys (*Macaca nemestrina*). *Vision Res* 16:517–521.
- Medsing A, Nischal KK (2015) Pediatric cataract: Challenges and future directions. *Clin Ophthalmol* 9:77–90.
- Lambert SR (2016) Changes in ocular growth after pediatric cataract surgery. *Dev Ophthalmol* 57:29–39.
- Ganesh S, et al. (2014) Results of late surgical intervention in children with early-onset bilateral cataracts. *Br J Ophthalmol* 98:1424–1428.
- Ma F, Ren M, Wang L, Wang Q, Guo J (2017) Visual outcomes of dense pediatric cataract surgery in eastern China. *PLoS One* 12:e0180166.
- Maurer D, Lewis TL, Brent HP, Levin AV (1999) Rapid improvement in the acuity of infants after visual input. *Science* 286:108–110.
- Jacobson SG, Mohindra I, Held R (1981) Development of visual acuity in infants with congenital cataracts. *Br J Ophthalmol* 65:727–735.
- Hartmann EE, Lynn MJ, Lambert SR; Infant Aphakia Treatment Study Group (2014) Baseline characteristics of the infant aphakia treatment study population: Predicting recognition acuity at 4.5 years of age. *Invest Ophthalmol Vis Sci* 56:388–395.
- Lesueur LC, Arné JL, Chapotot EC, Thouvenin D, Malecaze F (1998) Visual outcome after paediatric cataract surgery: Is age a major factor? *Br J Ophthalmol* 82:1022–1025.
- Latif K, Shakir M, Zafar S, Rizvi SF, Naz S (2015) Outcomes of congenital cataract surgery in a tertiary care hospital. *Pak J Ophthalmol* 30:28–32.
- Kwon M, Liu R, Chien L (2016) Compensation for blur requires increase in field of view and viewing time. *PLoS One* 11:e0162711.
- Smith FW, Schyns PG (2009) Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychol Sci* 20:1202–1208.
- Goffaux V, Hault B, Michel C, Vuong QC, Rossion B (2005) The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception* 34:77–86.
- Young AW, Hellawell D, Hay DC (1987) Configurational information in face perception. *Perception* 16:747–759.
- Peterson MA, Rhodes G (2003) *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes* (Oxford Univ Press, Oxford).
- Maurer D, Mondloch CJ, Lewis TL (2007) Sleeper effects. *Dev Sci* 10:40–47.
- Le Grand R, Mondloch CJ, Maurer D, Brent HP (2001) Neuroperception. Early visual experience and face processing. *Nature* 410:890.
- Le Grand R, Mondloch CJ, Maurer D, Brent HP (2004) Impairment in holistic face processing following early visual deprivation. *Psychol Sci* 15:762–768.
- Krizhevsky A, Sutskever I, Hinton GE (2012) *ImageNet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems (Curran Assoc, Red Hook, NY), pp 1097–1105.
- Ng HW, Winkler S (2014) A data-driven approach to cleaning large face datasets. *International Conference on Image Processing (ICIP)* (IEEE, Paris), pp 343–347.
- Kiorpes L, Movshon JA (1998) Peripheral and central factors limiting the development of contrast sensitivity in macaque monkeys. *Vision Res* 38:61–70.
- Yamins DLK, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci* 19:356–365.
- Cynader M, Berman N, Hein A (1976) Recovery of function in cat visual cortex following prolonged deprivation. *Exp Brain Res* 25:139–156.
- Fagioli M, Pizzorusso T, Berardi N, Domenici L, Maffei L (1994) Functional postnatal development of the rat primary visual cortex and the role of visual experience: Dark rearing and monocular deprivation. *Vision Res* 34:709–720.
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444.
- Swindale NV, Mitchell DE (1994) Comparison of receptive field properties of neurons in area 17 of normal and bilaterally amblyopic cats. *Exp Brain Res* 99:399–410.
- Witthoft N, et al. (2016) Reduced spatial integration in the ventral visual cortex underlies face recognition deficits in developmental prosopagnosia. *bioRxiv*:10.1101/051102. Preprint, posted April 29, 2016.
- Ellemberg D, Lewis TL, Maurer D, Brar S, Brent HP (2002) Better perception of global motion after monocular than after binocular deprivation. *Vision Res* 42:169–179.
- McKyton A, Ben-Zion I, Doron R, Zohary E (2015) The limits of shape recognition following late emergence from blindness. *Curr Biol* 25:2373–2378.
- Robbins RA, Nishimura M, Mondloch CJ, Lewis TL, Maurer D (2010) Deficits in sensitivity to spacing after early visual deprivation in humans: A comparison of human faces, monkey faces, and houses. *Dev Psychobiol* 52:775–781.
- Robbins RA, Shergill Y, Maurer D, Lewis TL (2011) Development of sensitivity to spacing versus feature changes in pictures of houses: Evidence for slow development of a general spacing detection mechanism? *J Exp Child Psychol* 109:371–382.
- Patel DE, et al.; OPTIC Study Group (2015) Study of optimal perimetric testing in children (OPTIC): Normative visual field values in children. *Ophthalmology* 122:1711–1717.
- Dobkins KR, Anderson CM, Kelly J (2001) Development of psychophysically-derived detection contours in L- and M-cone contrast space. *Vision Res* 41:1791–1807.
- Lewkowicz DJ (1996) Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform* 22:1094–1106.
- Colombo J (2001) The development of visual attention in infancy. *Annu Rev Psychol* 52:337–367.
- Parker DM, Lishman JR, Hughes J (1992) Temporal integration of spatially filtered visual images. *Perception* 21:147–160.
- Goffaux V, et al. (2011) From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cereb Cortex* 21:467–476.
- Wilson HR, Blake R, Halpern DL (1991) Coarse spatial scales constrain the range of binocular fusion on fine scales. *J Opt Soc Am A* 8:229–236.
- Rohaly AM, Wilson HR (1993) Nature of coarse-to-fine constraints on binocular fusion. *J Opt Soc Am A Opt Image Sci Vis* 10:2433–2441.
- Dominguez M, Jacobs RA (2003) Developmental constraints aid the acquisition of binocular disparity sensitivities. *Neural Comput* 15:161–182.
- Griffiths SK, Brown WS, Jr, Gerhardt KJ, Abrams RM, Morris RJ (1994) The perception of speech sounds recorded within the uterus of a pregnant sheep. *J Acoust Soc Am* 96:2055–2063.
- Ragó A, Honbolygó F, Róna Z, Beke A, Csépe V (2014) Effect of maturation on suprasegmental speech processing in full- and preterm infants: A mismatch negativity study. *Res Dev Disabil* 35:192–202.
- Gordon RA, Donzis PB (1985) Refractive development of the human eye. *Arch Ophthalmol* 103:785–789.
- Bailly C, O'Keefe M (2012) Paediatric aphakic glaucoma. *J Clin Exp Ophthalmol* 3:203.
- Sinha P, Held R (2012) Sight-restoration. *F1000 Med Rep* 4:17.
- Samaria F, Harter A (1994) Parameterisation of a stochastic model for human face identification. *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision* (IEEE, Sarasota, FL).
- Ringach DL (2002) Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol* 88:455–463.
- Niell CM, Stryker MP (2008) Highly selective receptive fields in mouse visual cortex. *J Neurosci* 28:7520–7536.