






Gaze following requires early visual experience

Ehud Zohary^{a,1,2}, Daniel Harari^{b,1} , Shimon Ullman^{b,1,2} , Itay Ben-Zion^c, Ravid Doron^d, Sara Attias^a, Yuval Porat^a, Asael Y. Sklar^a , and Ayelet Mckyton^{e,1}

Edited by David Heeger, New York University, New York, NY; received October 15, 2021; accepted March 3, 2022

Gaze understanding—a suggested precursor for understanding others' intentions—requires recovery of gaze direction from the observed person's head and eye position. This challenging computation is naturally acquired at infancy without explicit external guidance, but can it be learned later if vision is extremely poor throughout early childhood? We addressed this question by studying gaze following in Ethiopian patients with early bilateral congenital cataracts diagnosed and treated by us only at late childhood. This sight restoration provided a unique opportunity to directly address basic issues on the roles of “nature” and “nurture” in development, as it caused a selective perturbation to the natural process, eliminating some gaze-direction cues while leaving others still available. Following surgery, the patients' visual acuity typically improved substantially, allowing discrimination of pupil position in the eye. Yet, the patients failed to show eye gaze-following effects and fixated less than controls on the eyes—two spontaneous behaviors typically seen in controls. Our model for unsupervised learning of gaze direction explains how head-based gaze following can develop under severe image blur, resembling preoperative conditions. It also suggests why, despite acquiring sufficient resolution to extract eye position, automatic eye gaze following is not established after surgery due to lack of detailed early visual experience. We suggest that visual skills acquired in infancy in an unsupervised manner will be difficult or impossible to acquire when internal guidance is no longer available, even when sufficient image resolution for the task is restored. This creates fundamental barriers to spontaneous vision recovery following prolonged deprivation in early age.

joint attention | gaze | blind | cataract | vision

The direction of gaze of others is often an excellent cue for their immediate intentions and goals (1). Our ability to shift our gaze to the object of interest of another person develops in infancy and has been suggested as one of the first steps toward developing joint attention and a “theory of mind” (2). By 12 mo, infants follow another person's head direction more often when the person's eyes are open than when they are closed, indicating that they understand when others are “visually connected” to the external world (3). Gaze following reflects an understanding of the other person's point of view as well as the availability and likelihood of his future actions. Typically, our gaze direction is aligned with our focus of visual attention, and one's eye position is a better predictor of the object of interest and future actions than head or body orientation. Moreover, changes in eye position (e.g., focusing on an object) typically lead to head and body turns (4) and often enable reliable prediction of future hand actions toward a target object. It is no surprise, therefore, that adult humans follow eye gaze more reliably than head direction (5).

Gaze following is not limited to humans; macaque monkeys also respond faster to targets appearing in the direction of an actor's gaze (6). Gaze following is also not strictly reflexive, even in monkeys. It entails a deep understanding of social situations. For example, social status (e.g., dominance) affects gaze following in monkeys (7). Macaques, similar to humans, can also covertly attend to peripheral targets, allowing them to conceal their intent. Likewise, chimpanzees seem to know what conspecifics can and cannot see given their viewing perspective and use this knowledge to devise effective strategies in naturally occurring food competition situations (8).

From a computational standpoint, the early acquisition of gaze following is surprising, since the task is very difficult. Gaze following requires noticing fine details in the observed individual (i.e., the exact head orientation and eye position) and relating these to a direction in space, toward which that individual is gazing. Computationally, the natural development of understanding gaze direction is remarkable considering that this task (like most other visual tasks) is learned in an unsupervised manner, without explicit external guidance. Current computational models of vision rely on massive explicit supervision (9–11). Without it, learning is very limited (12). In particular, understanding gaze direction was noted as a difficult task without explicit supervision

Significance

Early in life, humans spontaneously learn to extract complex visual information without external guidance. Current vision models fail to replicate such learning, relying instead on extensive supervision. A classic example is gaze understanding, an early-learned skill useful for joint attention and social interaction. We studied gaze understanding in children who recovered from early-onset near-complete blindness through late cataract surgery. Following treatment, they acquired sufficient visual acuity for detailed pattern recognition, but they failed to develop automatic gaze following. Our computational modeling suggests that their learning is severely limited due to reduced availability of internal self-supervision mechanisms, which guide learning in normal development. The results have implications to understanding natural visual learning, potential rehabilitation, and obtaining unsupervised learning in vision models.

Author contributions: E.Z., D.H., S.U., S.A., Y.P., and A.M. designed research; E.Z., D.H., I.B.-Z., R.D., S.A., Y.P., and A.M. performed research; D.H., S.A., Y.P., A.Y.S., and A.M. analyzed data; and E.Z., D.H., S.U., A.Y.S., and A.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹E.Z., D.H., S.U., and A.M. contributed equally to this work.

²To whom correspondence may be addressed. Email: udiz@mail.huji.ac.il or shimon.ullman@gmail.com.

This article contains supporting information online at [http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2117184119/-DCSupplemental](https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2117184119/-DCSupplemental).

Published May 12, 2022.

(13–16). However, recently, Ullman et al. (17) showed that a series of unsupervised computational learning steps can lead to the understanding of gaze direction as well as hand recognition, mimicking normal visual development in human infants. Hand recognition is another extremely difficult computational task, which is quickly mastered by infants. Within 6 to 10 mo, infants expect hands to make contact with objects (18) and to cause them to move (19). At about the same time, infants often shift their gaze from faces to hands engaged in object manipulation (20, 21). Ullman et al. (17) suggested that these steps are exactly the necessary elements for hand recognition in early visual learning. The model utilized an internal teaching signal, called a “mover” event. A mover event is evoked when a moving element (most often a hand) comes into contact with a stationary object and causes it to move. This common and easily detectable event (e.g., seeing people manipulating objects) is repeatedly experienced by all infants. Crucially, it has proven to be a reliable indicator of a hand making contact with an object, thereby facilitating hand recognition by the model. The model also explains how infants learn to recognize and follow the direction of gaze of others. It utilizes the fact that people typically gaze directly at the object they manipulate, particularly just before and during initial object contact. The model detects hand–object contact (i.e., mover events) and extracts the corresponding face images at that time. This information allows the model to learn to associate the face appearance in the image, including head orientation and pupil position, with a specific vector direction from the head to the object of contact. This association leads to a fast and reliable understanding of the direction of gaze of others (17).

Humans are masters in predicting others’ intentions from fine visual signals, but this level of sophistication develops slowly over time. Obviously, the viewer must have sufficient resolution to correctly register the head orientation and the eyes’ position in their orbits to obtain head and eye gaze understanding, respectively. A coarse ability to distinguish between direct and averted gaze is present already in newborns (22), probably because direct gaze is associated with parental contact and safety, but fine-grained assessment of direct gaze reaches maturation only many years later (23). Gaze following and gaze understanding are primarily relevant to social interactions and require prolonged learning after attaining sufficient visual acuity. At 1 mo of age, normally developing infants can detect grating patterns of ~ 1 to 2 cycles per degree (cpd), enough to tell head orientation but not eye position from 1 m. Still, they do not follow the head direction of others until 3 to 6 mo of age, when their visual acuity is ~ 4 to 8 cpd (24–26). Interestingly, at 1 to 2 mo, babies change their eye-scanning pattern from the viewed face perimeter to its internal features [e.g., the eyes (27)]. Focusing on and attending the eyes is required for establishing eye gaze–following behavior. If the actor’s eyes are in motion (e.g., refocusing on a new object), attention is spontaneously captured by motion, and eye gaze following occurs already at 2 to 4 mo (25). At 3 to 6 mo, infants gain sufficient visual acuity to tell another person’s eye position from ~ 1 m, but when the actor’s eyes are static, eye gaze following emerges only at 1 y when visual acuity is well above 10 cpd. Thus, being able to detect the relevant cue does not entail its immediate use for gaze following. Further visual experience is required to associate gaze direction with the target of gaze and acquire predictive and social behavior accordingly.

However, what if a growing child had extremely poor visual acuity in the early period, when gaze direction is normally attained? Would improvement in visual acuity following

cataract surgery (21) (allowing for eye position extraction) be sufficient for recapitulating the normal developmental process, even at late childhood (28)?

We had a chance to study this issue in an exceptional group of children that suffered from dense bilateral cataract since early infancy, which were found and surgically treated by our team only years later. We studied their gaze following by assessing the cueing effect in a gaze-cueing task (29), a task that produces automatic orienting of attention in response to an actor’s averted gaze direction. We found that, following cataract surgery, our late-treated participants were able to orient attention to the object of observed gaze when the cue was provided by head orientation, but they failed to do so when the cue was provided by eye position. Furthermore, although these late-treated patients naturally focused on faces in an image, they tended to fixate less on the person’s eyes or on the object the person was looking at. This population provides unique testing conditions for our theoretical approach because the visual information required for head-based gaze extraction is available prior to the medical intervention, while for eye-based gaze extraction, the relevant information becomes available only following the treatment.

We compare these results to those obtained by the computational model [of Ullman et al. (17)] under extremely reduced vision conditions (mimicking the blur experienced by the patients prior to surgery) using low-resolution images and better-resolution images (reflecting their condition after surgery). Fig. 1 depicts the basic capacities required in the model to deduce and follow head and eye gaze direction to the target object (Fig. 1 *A* and *B*, respectively). These capacities include three components: 1) detecting a mover event, 2) directing attention to the face/eyes, and 3) perceiving the head orientation/eye position in their orbits at the time of the mover event. A table indicating the above capacities for head, or eye gaze following and their availability in the late-treated patients is indicated in Fig. 1 *C* and *D*, respectively.

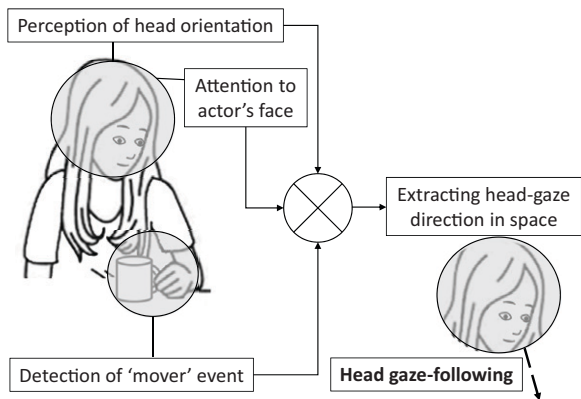
Results

1. Behavioral Experiments.

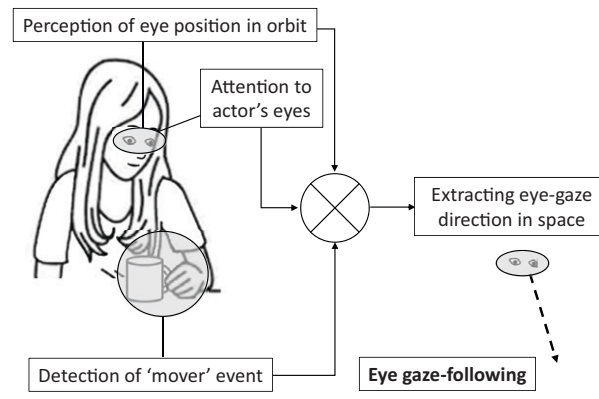
1.1 Visual acuity. We studied 19 Ethiopian children (age 12.7 ± 3.4 y; *SI Appendix*, Table 1) (30) who suffered from early-onset bilateral cataracts and were diagnosed and operated by our group many years later (at age 11.3 ± 3.3 y; Fig. 2*A*, yellow symbols). They participated in at least one of the behavioral experiments described below (sections 1.2 and 1.3). The contrast sensitivity function (CSF) was assessed on the same day as the behavioral experiments and in 17/19 cases, also before surgery (Fig. 2*B* and *SI Appendix*, Fig. 1). Typically, following surgery, the children showed dramatic improvement in their visual acuity (Fig. 2*C*; average preoperative cutoff frequency: 1.1 cpd; average postoperative cutoff frequency: 6.1 cpd; see also refs. 31 and 32), although this was still very poor with respect to normal visual acuity (33). The study also included 11 Israeli children (age 10.1 ± 3.3 y) with congenital cataracts that were surgically treated within a few months after birth (Fig. 2*A*, blue symbols). These children typically have only a slight loss of visual acuity (34, 35). A group of 88 children (age 8.8 ± 2.3 y) with typically developed vision served as controls.

1.2 Gaze-cueing experiment. We used the gaze-cueing paradigm in two experiments (Fig. 3*A*) and measured the effect of a gaze cue generated by a change in the eye position or head orientation. In both experiments, at the beginning of the trial, the participants were required to touch the nose of the seen face, thereby ensuring that they orient their attention to the face and

A Head gaze-following development



B Eye gaze-following development



C

Capacity	Our Findings			
	Pre Op (high blur)	Post Op (low blur)	Section in text Behavior	Models
Perception of head orientation	✓	✓	1.1	2.2
Detection of 'mover' event	✓	✓	---	2.1
Attention to actor's face	---	✓	1.4	---
Extracting head gaze direction	✓	✓	1.2	2.1
Head gaze-following	✓	✓	1.2	---
Head gaze understanding	---	✓	1.4	---
	---	✓	1.3	---

D

Capacity	Our Findings			
	Pre Op (high blur)	Post Op (low blur)	Section in text Behavior	Models
Perception of eye position in orbit	x	✓	1.1	2.2
Detection of 'mover' event	✓	✓	---	2.1
Attention to actor's eyes	---	x	1.4	---
Extracting eye gaze direction	x	x	1.2	---
Eye gaze-following	x	x	1.2	---
Eye gaze understanding	---	x	1.4	---
	---	x	1.3	---

Fig. 1. Gaze following development: modeling and findings. (A and B) A diagram describing the model's necessary requirements for developing head (A) and eye (B) gaze following by observing the actions of another person. In the congenital cataract patients, prior to surgery, the conditions described in A are available, but those in B are not. After the operation, the conditions depicted in both A and B are available, but despite this, eye gaze following is not established. (C and D) Our findings show the capacities for each task during the preoperative and postoperative stage for head (C) and eye (D) gaze following. The behavioral and model results indicate that the self-teaching mechanism for gaze following is unavailable beyond early development. V denotes an intact capacity, X denotes a deficit, and — denotes an untested capacity; preop, preoperative; postop, postoperative.

fixate near the actor's eyes ($\sim 2^\circ$) without obstructing the eyes themselves. Touching the nose led to a presentation of the gaze cue (eyes or head shifting to the left or right, perceived as apparent motion). Then, 300 ms later, a balloon appeared either in agreement with the gaze direction ("compatible") or on the opposite side ("incompatible"). Participants were instructed to touch the balloon as quickly as possible, and reaction time was monitored. Only patients that recognized our stimuli as faces, could locate the eyes in the face, and distinguished between static eyes and moving eyes in preliminary inclusions tests (*Methods* and *SI Appendix*, Fig. 2) participated in the gaze-cueing experiments (all early treated and 15 late treated). Control participants performed the experiments using highly blurred images (see Fig. 3B) to control for the poorer visual acuity of the late-treated patients (cutoff frequency of 1.6 cpd, worse than the blur experienced by any patient after surgery, Fig. 2C).

Fig. 3C shows the cue compatibility effect—the difference in response time (RT) between incompatible and compatible trials—for the different groups in the two experiments. To assess this quantitatively, we performed a repeated measures 3×2 ANOVA, with group (controls, early treated, and late treated) and experiment (eye and head direction) as the main factors. There was a significant difference between the groups ($F[2,137] = 3.6$; $P = 0.031$) and a highly significant interaction term ($F[2,137] = 13.1$; $P < 0.0001$; full ANOVA results are in *SI Appendix*, Table 2), confirming that the groups behaved differently in the two cueing conditions. The only significant differences between groups in either head- or eye-cueing effects were for the eye-cueing effect: controls versus late treated ($t[59] = 7.6$; $P < 0.0001$) and early versus late treated

($t[24] = 5.2$; $P < 0.0001$; Fig. 3C). When tested separately in each group, a head gaze compatibility effect was evident in all groups (late treated: M [mean] = 60 ± 53 ms, $t[13] = 4.2$, and $P = 0.001$; early treated: $M = 30 \pm 4$ ms, $t[10] = 2.3$, and $P = 0.048$; controls: $M = 42 \pm 34$ ms, $t[45] = 8.3$, and $P < 0.0001$). In contrast, an eye gaze cue compatibility effect was evident only in controls ($M = 52 \pm 20$ ms, $t[45] = 17.6$, and $P < 0.0001$) and in the early-treated ($M = 40 \pm 17$ ms, $t[10] = 7.8$, and $P < 0.0001$), but not in the late-treated ($M = 1.5 \pm 29$ ms, $t[14] = 0.2$, and $P = 0.84$), group. Individual data in each group of patients and the dependence of the cueing effect on various factors (visual acuity, time since surgery, etc.) are shown in *SI Appendix*, Fig. 3. In addition to the treated patients, we also tested three preoperative patients. These preoperative patients showed a dramatic gaze effect for head orientation ($M = 92 \pm 58$ ms) but no effect for an eye-directed gaze ($M = 5 \pm 46$ ms). While these results are obviously insufficient for making definite conclusions, they suggest that head gaze-following behavior is acquired by the late-treated patients prior to surgery despite the extreme image blur that they experienced.

An alternative interpretation might suggest that the head gaze-following behavior is acquired, at least in part, after the operation. We find this suggestion unlikely for the following reasons. First, as described in detail below, our simulations and model (section 2.1) show that the preoperative, low-resolution conditions provide sufficient visual information for learning head gaze following. In particular, we show that the learning process used for extracting head and eye gaze direction in normal conditions (17) can learn to extract head direction even under the extreme blur conditions mimicking those experienced by the

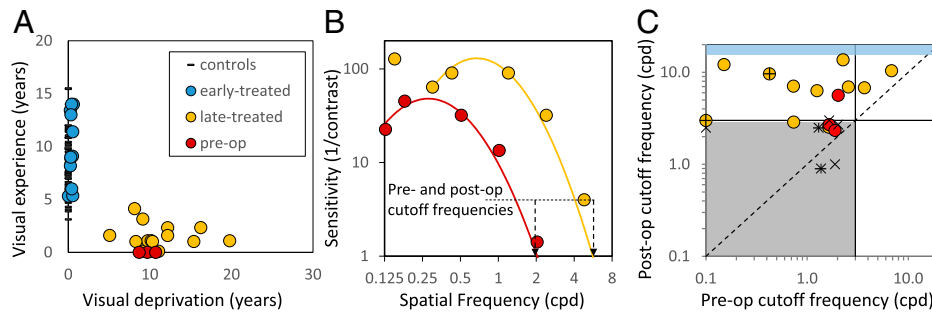


Fig. 2. Visual development and spatial acuity of the participants. (A) A scatter plot depicting the individual participants' visual development. The abscissa denotes the duration of visual deprivation in years (i.e., in the cataract-treated groups, the age at surgery; in controls, 0). The ordinate denotes the years of visual experience (i.e., in the cataract-treated groups, time since surgery; in controls, the age of the participant at testing). (B) The CSF of one late-treated participant tested both before (red circles) and after (yellow circles) surgery. The results illustrate characteristic improvement of spatial vision following surgery. The cutoff frequency is defined as the crossing point of the inverse parabola fitted to the data with the abscissa. (C) Scatter plot showing the post-operative cutoff frequency of the patients at the date of gaze-cueing testing as a function of their preoperative visual acuity. Yellow and red circles indicate late-treated patients who performed the gaze-cueing test after surgery and before surgery, respectively. Late-treated patients ($n = 6$) that did not pass the inclusion criteria for the gaze-cueing task are depicted by \times . Four late-treated patients (denoted by a superimposed $+$ sign) did not do the CSF test before surgery, and thus their preoperative visual acuity denoted here is their first CSF test result after surgery (<1 mo after surgery). The visual acuity of the early-treated patients ($n = 11$) was not assessed prior to surgery. Their postoperative cutoff frequency at the time of gaze-cueing testing was always better than the maximum spatial frequency. Their acuity is therefore depicted by the light blue region (above 13.6 cpd). The cutoff frequency for legal blindness (3 cpd according to NIH guidelines) is highlighted by a gray background square. Note that most late-treated patients were legally blind before surgery, but their visual acuity improved substantially after surgery, such that they were no longer considered legally blind.

late-treated patients prior to surgery. Second, the alternative explanation requires an additional assumption, namely that the restored vision is effective for head, but not eye, following. In contrast, both the empirical results and computational simulations show that the available visual information under restored conditions is sufficient for extracting both head and eye orientation. Third, our preoperative data, although limited, are consistent with our proposed interpretation.

The results also show the value of early compared to late surgical intervention. Unlike the late-treated participants, our early-treated participants showed a clear eye gaze cueing effect similar to that found in controls. Thus, gaze-following mechanisms based on eye position information develop normally despite a brief deprivation period in early development (typically 4 to 6 mo).

1.3 Explicit gaze understanding. We also tested 10 of the late-treated participants (which previously did the eye gaze-cueing experiment) on explicit gaze understanding by reporting the gazed-upon object. As described previously, the experiment started with presentation of a face changing its head or eye gaze direction to one of *two* balloons appearing on *both* sides.

Participants were asked to touch the balloon that the actor was gazing at. All 10 participants succeeded significantly above chance level ($P < 0.05$) in the head gaze condition, but only 3 succeeded in the eye gaze condition (SI Appendix, Supplementary Text 2), while the rest were at chance level. Controls did the test flawlessly despite the blur imposed. The results indicate that again, as a group, the late-treated participants fail to generate an association between gaze and the object of gaze.

1.4 Eye movement patterns under free-viewing conditions. We found that late-treated patients failed to automatically use an eye gaze cue to respond faster to a target presented at the cued location. On the other hand, they were able to use a head orientation cue to generate a speedier response to the head-directed target. Is similar behavior observed in the *eye movement patterns* of our late-treated participants under free-viewing conditions?

In the first test, "gazed object" (Fig. 4A), participants viewed (on the screen) a person shifting his eye or head orientation to the left or to the right toward one of two identical objects. The hypothesis was that if this cue was effective, the actor's gaze direction would lead to longer fixations on the cued object than on the noncued object, captured by a positive-cued object

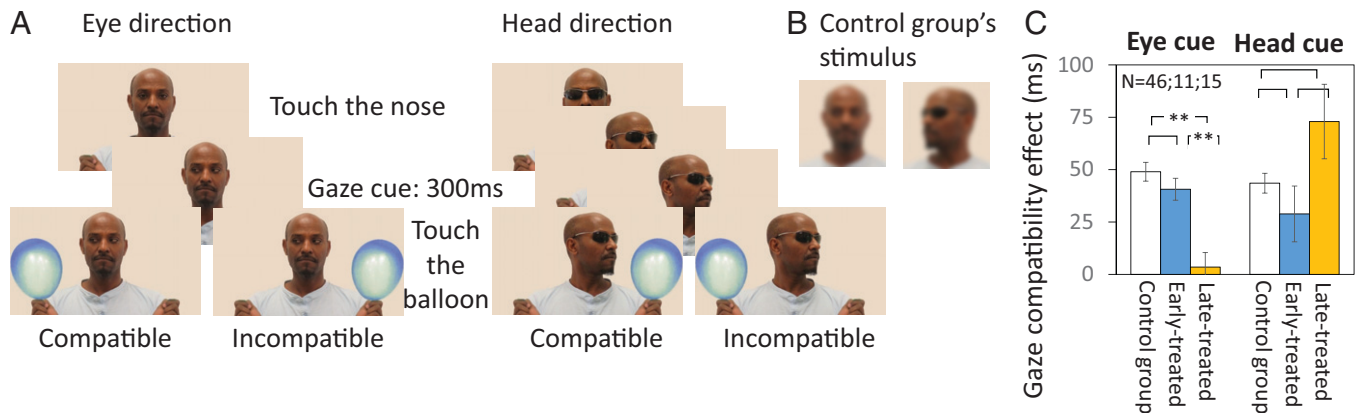


Fig. 3. Gaze-cueing experiment stimuli, design, and group results. (A) Experimental design of the main gaze-cueing experiments, testing compatibility effect to eye (Left) and head (Right) gaze cues. (B) Example of blurred stimuli seen by controls. (C) Group results for the eye (Left) and the head (Right) direction experiments, depicting the group average cue-compatibility effect (RT of incompatible minus compatible trials) in the control (white), the early-treated (blue), and the late-treated (yellow) groups. Error bars denote SEM. The numbers of participants from each group in the two experiments (N) are indicated at the top. Horizontal bars indicate direct comparisons between group effects. Two asterisks (**) denote statistically significant differences; $P < 0.001$.

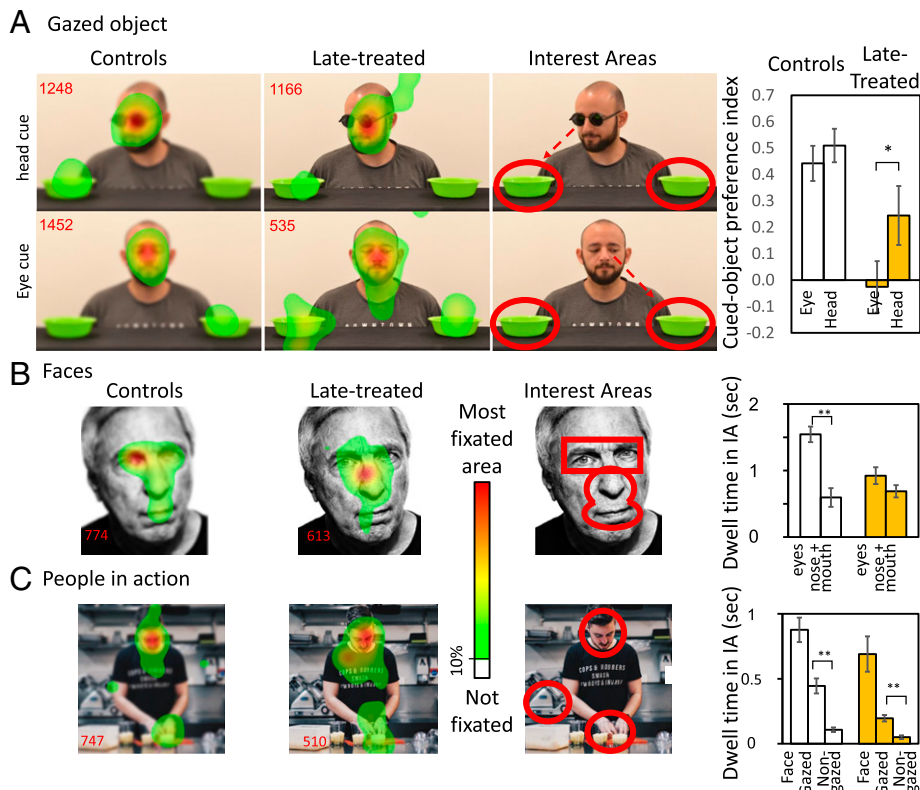


Fig. 4. Eye movement patterns during free viewing. (A) Fixation maps of controls (left column; $n = 31$) and late-treated participants (middle column; $n = 9$) during observation of an actor gazing at a target object (depicted, for illustration only, by a red arrow) indicated by head orientation (upper row) or eye position (lower row). Predefined interest areas are depicted by red ellipses (right column). The numbers on each image denote the maximum time spent fixating on a specific position in the image (group average smoothed with a Gaussian kernel of 1°). (Right) Bar plots depicting the mean cued-object preference index [(cue congruent - incongruent]/[congruent + incongruent] fixation dwell times) for the control (white) and the late-treated (yellow) groups. Positive values indicate a fixation preference for the cued object. (B and C) Fixation maps of controls (first column; $n = 11$) and late-treated participants (second column; $n = 9$) when observing an image of an exemplary face (B) and people in action (C), respectively. (Right) Bar plots depict the cumulative duration of fixations (dwell time) in each interest area (IA) for the two groups. An asterisk (*) denotes $P < 0.05$; **, $P < 0.005$. In all tests, controls viewed a blurred version of the images (smoothed with a Gaussian kernel). Error bars depict SEM.

preference index. Indeed, this was the case in controls: both head orientation ($t[30] = 8.05$; $P < 0.001$) and viewed eye position ($t[30] = 6.63$; $P < 0.001$) were effective cues with no significant difference between the two conditions ($t[30] = 1.34$; $P = 0.190$; Fig. 4A, white bars). In the late-treated participants, head orientation was marginal (likely due to lack of statistical power: $t[8] = 2.17$, $P = 0.060$), while eye position was totally ineffective ($t[8] = -0.26$; $P = 0.801$) in drawing one's gaze toward the cued object. Indeed, unlike controls, there was a significant difference between the two conditions ($t[8] = 2.41$; $P = 0.042$; Fig. 4A, yellow bars). This lends further support to our conclusion that the late-treated patients do not utilize eye gaze information to direct their own gaze to other actors' objects of gaze.

Next, to test whether the late-treated participants preferentially attend to the eyes of another person, a prerequisite for acquiring eye gaze following, we presented participants with images of faces. Prior to the experiment, interest areas were defined for the "eyes" and for the "nose-and-mouth" regions (Fig. 4B). There were significant differences between the groups ($F[1,18] = 8.054$; $P = 0.011$) and areas ($F[1,18] = 15.873$; $P < 0.001$) as well as, critically, a significant interaction term ($F[1,18] = 5.783$; $P = 0.027$). While control participants fixated longer on the eyes in comparison to the nose-and-mouth region ($t[10] = 4.2$; $P = 0.0017$; Fig. 4B), the late-treated patients did not ($t[8] = 1.2$; $P = 0.23$).

The two-image scanning behaviors expressed by the late-treated patients may stem, at least to some degree, from a common source: the actor's eyes attract less attention than typical

vision-developing controls. Attention to the actor's eyes is probably necessary to effectively capture the seen eye position when the hand establishes contact with the object (i.e., the mover event). These two elements (eye and target object position) must be associated to generate a reliable representation of eye gaze directions. Without them, the model will fail to learn eye-based gaze following, and, obviously, there will be no clear eye position cueing effect (Fig. 1).

Last, we had our participants view natural images depicting an object-related action in which the eye and head direction were in accordance with the hand manipulating the object. These "people in action" pictures (Fig. 4C) had three interest areas: 1) the face, 2) the gazed-upon object, and 3) a nongazed object in the image (different object of similar size). Both late-treated ($t[8] = 5.9$; $P = 0.0004$) and control ($t[10] = 5.1$; $P = 0.0002$) groups fixated longer on the gazed object than on the nongazed object. Thus, when seeing images showing typical object manipulation scenes, in which people orient both their head and eye gaze directly at the objects, patients show clear preference for the target object over other regions of the image. This is consistent with our behavioral experiment, showing that the late-treated patients show spontaneous gaze following in response to head orientation cues.

Could the above results be due to the nystagmus experienced by all our late-treated participants? A recent paper (36) showed that late-treated cataract patients are able to execute accurate visually-guided eye movements to targets despite having severe nystagmus. This was also apparent in our case; patients

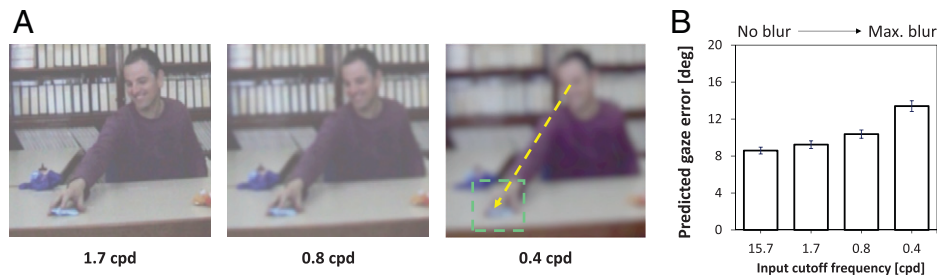


Fig. 5. Development of gaze following in various blur conditions. (A) Examples of blurred images at various cutoff frequencies (1.7, 0.8, and 0.4 cpd). A mover event (green box) is clearly detectable even at the largest blur (0.4 cpd) in the dynamic sequence, although the hand is difficult to recognize by its own appearance. The gaze direction (yellow arrow) can be interpretable using the head orientation but not the eyes' gaze. (B) Prediction of gaze direction. The angular error in degrees (deg) between the true and predicted gaze directions of a neural network (Resnet50). Error bars denote SEM. The neural network was trained to predict gaze-direction angle from face images under increasing input blur levels. Mover event locations were used as gaze target positions. Gaze directions were in the range of (-90° , $+90^\circ$) left to right, respectively. Input images were blurred using Gaussian filters with a cutoff frequency of 15.7 (no blur), 1.7, 0.8, and 0.4 cpd (corresponding to kernel spatial SD of 0, 8, 16, and 32 pixels, respectively).

typically fixated on the face before making a second eye movement landing in the vicinity of the target object. However, as expected, nystagmus indeed resulted in broader scanning patterns in the patients than in controls (e.g., broader "heat maps"; Fig. 4 A and C). Thus, due to the uncontrolled jitter in eye position, the absolute dwell time in each interest area was typically shorter in the patients than in controls. Note, however, that our conclusions from the eye movement analysis are centered on the comparison between the dwell times in the various interest areas *within* the patient population. These comparative measures are clearly unaffected by the nystagmus.

2. Computational Experiments. To gain a wider interpretation of the behavioral findings, we studied the computational requirements for learning to extract gaze direction as a developmental (unsupervised) process. In particular, we were interested in studying the implications of image degradation at various blur levels and a partial recovery of visual acuity (similar to that experienced by our patients) on the possibility of learning gaze direction based on head or eye orientation cues.

2.1 Unsupervised learning of gaze following under blur conditions.

We used a computational model for unsupervised learning of gaze direction, which follows the order of developmental steps in which human infants acquire the ability to follow the gaze of others (17). Specifically, we studied the consequences of highly degraded visual input to the model's performance. To that end, we assessed the detection level of mover events when the input was limited to low spatial frequencies, similar to the extreme visual acuity limitations imposed by cataract prior to surgery (and to a lesser degree also after surgery). The model's visual inputs were video sequences showing people moving objects on a table. The videos were blurred to various levels (Fig. 5A, Methods, and SI Appendix, Supplementary Text 1). The model's ability to detect a mover event, assessing detection precision (i.e., the fraction of true hand-object contacts out of all the detected mover events) was above 60% (SI Appendix, Fig. 5), even at the highest blur level regimen (corresponding to a cutoff frequency of 0.4 cpd). Crucially, this precision, although far from perfect, proved to be a sufficiently reliable teaching signal for learning the gaze direction of others; the model's performance was similar to human level (37) even at the highest blur condition (Fig. 5B and Methods). We conclude that mover event detection is a reliable teaching signal for gaze following even at poor visual acuity conditions similar or worse than that experienced by the late-treated patients prior to surgery.

2.2 Spontaneously developed face representations in restored acuity conditions can discriminate eye positions.

The computational simulations above show that gaze direction can be acquired spontaneously even under extreme blur conditions, in which the head orientation contributes the most, and the contribution of the eyes' orientation is likely to be minor at best due to the low resolution. In the next computational test, we wanted to ascertain that under the postoperative conditions, there is sufficient resolution to discriminate eye orientations. We assume that following surgery, patients are unlikely to get specific training for eye position but gain experience with more general face-related tasks, such as face identification. Therefore, we used activation levels of intermediate layers in a convolutional neural network trained for face identification. These activation levels serve as face representations used in the network to identify faces and encode information on facial features, including the eyes. Our simulations tested if these face representations have sufficient information to discriminate between eye positions under different blur conditions. Note that the network was not explicitly trained to perform this discrimination but rather to identify faces, and, therefore, we regard the network's internal representations as analogous to the spontaneously developed representations in the patients. As a baseline, we first performed this evaluation procedure (learning face representations and using them to discriminate head and eye orientations) using a neural network which was trained using images at normal resolution (e.g., a resolution corresponding to a cutoff frequency of 17.6 cpd; Fig. 6C). Its performance level is depicted in Fig. 6D and E (gray bars).

In simulating preoperative acuity conditions, we applied a different training regime, under high-blur conditions (using a Gaussian filter with a cutoff frequency of 0.8 cpd; Fig. 6A), similar to the patients' preoperative acuity. Under this training regime, the network was able to reliably discriminate between left and right head orientations (Fig. 6D, black bar at 0.8 cpd; $M = 88\%$; $SEM = 3\%$) but was unable to discriminate between left and right eye positions (Fig. 6E, black bar at 0.8 cpd; $M = 51\%$; $SEM = 1\%$). Moreover, the network's internal representation was unable to recover eye position even when exposed (with no further training) to images at higher resolution (Fig. 6E, black bar at 3.3 cpd; $M = 54\%$; $SEM = 1\%$). However, when we applied additional training to the network (on the task of face identification) with input at higher resolution (using a Gaussian filter with cutoff frequency of 3.3 cpd), similar to the patients' postoperative acuity (Fig. 6B), the network yielded improved discrimination ability between the eye

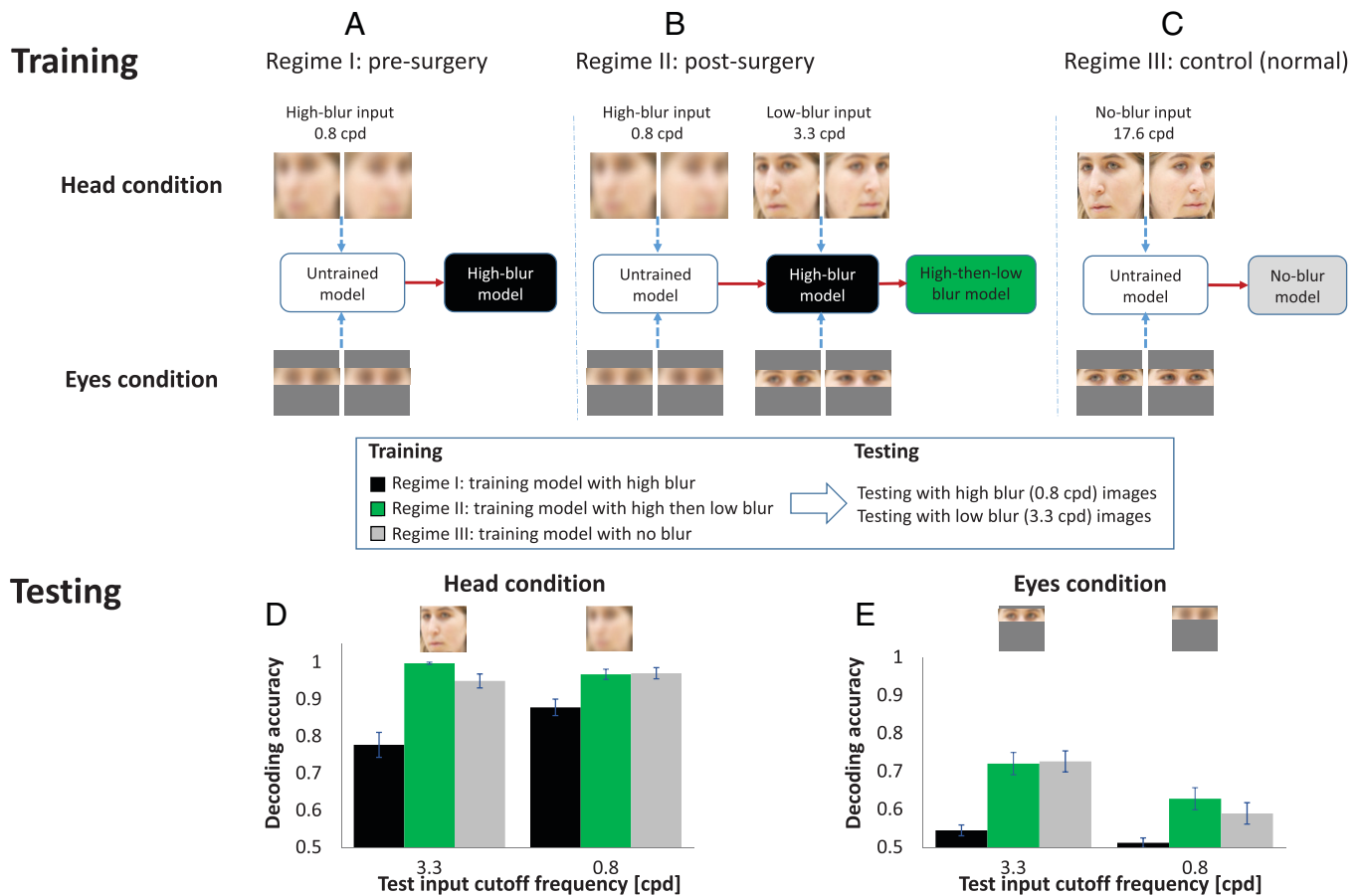


Fig. 6. Computational discrimination between faces looking left or right. Model discrimination was based on the activation of intermediate layers of the network (Resnet50). (A–C) Training. Three networks were trained to identify faces using images of either full faces (head condition) or only the eyes region (eye condition) at various blur levels. Blue arrows indicate the input to each network, and red arrows indicate the network’s development (phases of training) in time. (A) Regimen I: “presurgery” (in black). Training on images at a high-blur level (using a Gaussian filter with cutoff frequency of 0.8 cpd) similar or worse than preoperative conditions. (B) Regimen II: “postsurgery” (in green). Training first on images at a high-blur level (mimicking preoperative vision; cutoff frequency of 0.8 cpd) and then further training on images at low-blur levels (cutoff at 3.3 cpd) similar or worse than the postoperative visual acuity. (C) Regimen III: “control (normal)” (in gray). Training on images at the highest resolution with no blur (cutoff at 17.6 cpd). (D and E) Testing. The networks under the three training regimens were evaluated for left/right discrimination of head orientations (D) or eye directions (E), as seen in B. Bar colors correspond to the three regimens (black, high blur; green, high then low blur; gray, no blur). Note that head orientation/eye direction were not explicitly learned during training (to identify faces). The input for testing was either highly or moderately blurred (0.8 or 3.3 cpd, respectively). Error bars represent SE over test images. Chance level is 50%.

positions (Fig. 6E, green bars; $M = 72\%$; $SEM = 3\%$). This result is in line with the behavioral results of our postoperative patients, confirming that computationally, following surgery, the patients’ spontaneously-developed face representations are sufficient to reliably discriminate the position of the eyes in their orbit. Nevertheless, despite this newly acquired ability, the patients did not use this information to develop eye gaze-following behavior.

Discussion

Our empirical results show that, similar to controls, the early-treated patients spontaneously followed the gaze of others when gaze direction was indicated by either head orientation or eye position cues. In contrast, the late-treated patients showed a spontaneous cue compatibility effect only when gaze was indicated by the head direction of others. Our paper’s experimental aim was to study whether one develops automatic eye gaze following if the cue becomes available only in late childhood. Thus, only patients with a minimal spatial resolution, who could reliably perceive the eye position cue took part in our study. Naturally, this population typically had better visual acuity than in other studies of late-treated patients (21, 31, 34, 38). However, despite this, our late-treated patients were

unable to automatically follow the eye gaze of others. Moreover, their eye movement patterns during free viewing of gaze-cueing movies matched the above distinction; they displayed normal head- but not eye-based target tracking.

Our computational results show that under extreme image blur, it is possible to discriminate between head, but not between eye, orientations. However, postoperative resolution is sufficient for distinguishing the eye orientation. With respect to the unsupervised learning of gaze direction, we found that our self-teaching model can correctly identify mover events under severe low-pass-filtering conditions, as typically experienced by the preoperative patients. We further showed that despite this severe image blur, the model can effectively utilize these mover events to acquire gaze-following behavior based on head orientation cues but not eye orientation. Our model is therefore consistent with the results of the three preoperated participants and with the hypothesis that head gaze following was established already prior to surgery where both head orientation discrimination and unsupervised learning of gaze direction were present, while eye gaze following was not. Empirically, eye gaze following was not acquired even after surgery despite the fact that eye orientation discrimination was acquired by the late-treated patients.

Note, however, that even a massively blurred version of the face and eyes, worse than that experienced by the late-treated patients, was enough to induce automatic eye gaze following in the control group. Apparently, poor image resolution in itself is not a limitation for eye gaze following if one had previous early visual experience (including finer spatial frequencies) to establish the association (e.g., direction vector) between viewed eye position and the target location in the world.

The late-treated participants have apparently become very proficient using information from head pose (e.g., low spatial frequencies) to guide their gaze-tracking behavior. One might posit that head orientation is sufficient for efficient gaze following, and the late-treated participants simply lack motivation to acquire eye gaze following. However, recently, Harari and colleagues (37) specifically addressed the accuracy of gaze estimation in natural conditions when only the eye region was visible versus when the face without the eyes was visible. They found that the eye region yielded a more accurate estimation of gaze direction than the face-without-eyes condition. Thus, in line with common intuition, the eyes provide fine information that improves our estimate of the gaze direction of others. Indeed, early in development (at age 3 to 6 mo), infants rely on head orientation to judge others' gaze direction and follow it. However, with further improvement of visual acuity, at above 1 y of age, they refine their behavior by including the eye position in its orbit as well as head orientation to assess gaze direction. Thus, the internal teaching process guiding the eye-based gaze-direction extraction is still available after head-based gaze direction has been learned. Indeed, the early-treated patients, who regained their vision well before 1 y of age, were able to acquire eye gaze-following behavior.

However, the situation was different in our late-treated patients. We suggest that the acquisition of useful visual functions following surgery may fail even when pattern identification is restored, because the unsupervised learning process is not available anymore. The deficit may be in the failure to register the eyes' pattern and use this cue to guide gaze following. Specifically, for effective usage of mover events to establish eye gaze following, the actor's eye position must be registered (by refixation on the eyes or covert attention) and associated with his hand position around the time of initial object movement to guide the gaze-following learning. In our testing, the late-treated patients tended to fixate less on the eyes and gaze-target objects than controls, indicating that they assign less overt attention (and significance) to these elements in the image.

Interestingly, Senju et al. (39) found that infants of blind parents show a similar pattern: they look less at an adult's eyes than the mouth region, and gaze less at the object to which the adult is looking. Their deficiencies in eye gaze processing become more pronounced at 12 mo of age. These observations are in line with our model's expectations. Since typically in the blind, their eye gaze is not aligned with the object of hand manipulation at the time of contact, the infant is less likely to establish veridical eye gaze following. Note that eye gaze following could possibly be learned later in life: If the internal guiding mechanism is still operative beyond infancy, babies will have ample opportunities to learn from interactions with people other than their blind parents. Further research may clarify this important point.

Could the eye gaze-cueing deficits reflect a side effect of some larger change in face processing? After all, holistic face processing is disrupted by early visual deprivation (40). It also develops during infancy on a similar trajectory as eye gaze following. Previous studies repeatedly found that even early-

treated cataract patients suffer from impaired holistic face processing (40, 41). However, in our study, early-treated patients showed eye gaze-cueing effects as large as the control group. Thus, at least mild deficiencies in holistic face processing, as reported previously (40), do not necessarily translate to a deficiency in eye gaze cueing. Still, it is conceivable that greater deficits in face processing, due to a much longer deprivation period in the late-treated participants, might result in a lack of eye gaze cueing among other problems in face processing. Testing this possibility, we found that, in controls, the eye gaze-cueing effect is as potent when faces are inverted (thereby disrupting holistic face processing) as when the faces are upright (*SI Appendix, Supplementary Text 5*). Thus, the lack of eye gaze-cueing effects in the late-treated participants is probably not simply a side effect of some larger deficit in holistic face processing.

To summarize, a substantial improvement of visual acuity following late surgery, sufficient for the detection of certain patterns and events that were previously unavailable, does not necessarily lead to the spontaneous acquisition of some high-level visual capacities and useful visual behavior. Interestingly, a similar lack of spontaneous acquisition years after surgical treatment was found in such patients for automatic imitation (42). Viewing a hand action performed by another person did not facilitate a faster response-compatible action and slower response-incompatible response as large as in typically developing peers. Our present results suggest that a long deprivation period precludes spontaneous acquisition of gaze following. This period is likely to be longer than 1 y, the period at which the early-treated participants had their surgery.

The early-treated group had longer visual experience than the late-treated group, so we cannot completely rule out the possibility that eye gaze following may develop spontaneously in the late-treated group with a sufficiently extended period of visual experience, but this appears unlikely. Our late-treated patients often had years of visual experience following surgery before testing. Moreover, a previous case study of a congenital cataract patient, surgically treated at age 12 and tested 20 y later (43), showed that she did not use eye position cues for gaze direction estimation and relied solely on head orientation. In contrast, infants with normal vision typically fully develop these capacities spontaneously within the first 2 y of age (26, 44, 45).

It is difficult to distinguish between two alternatives: 1) that the internal guiding mechanism is only available at an early stage in development, or 2) that the internal guiding mechanism still exists, but due to reduced motivation or other factors, it is no longer automatically utilized to guide eye-based gaze extraction. Indeed, head and eye direction are typically in tight alignment. Thus, head direction may be sufficient in most cases to indicate others' gaze direction.

Note, however, that in developing high-accuracy gaze following, a limitation of the period during which the mover-based guidance operates (e.g., option 1 above) may in fact be useful. The reason is that the teaching signal, provided by mover events, is inherently noisy, since some detected mover events are in fact not true hand-object interactions [e.g., a moving object hitting another object and causing it to move would be also registered as a mover event (17)]. Following the initial gaze learning, more accurate cues become available for learning. For example, after effective hand recognition has evolved, using actual hand-object contact will be more accurate than the original mover events. In general, when better cues become available, reducing the noisier and less-accurate cues can make the learning more effective.

Under the model assumptions of an early teaching signal with a critical period for acquisition, early-treated cataract patients may regain most functions after surgery. For late-treated patients, the model suggests three main categories of reduced preoperative vision in terms of their implications for postsurgery recovery: 1) extreme blur conditions, where mover events are not detectable (*SI Appendix, Supplementary Text 4*), and the resolution is sufficient only to discriminate limited head orientations (e.g., frontal from side view) but is insufficient to discriminate between any eye orientations (the model predicts that automatic postsurgery gaze following behavior from both head and eye orientations will not emerge); 2) high blur conditions, where resolution is sufficient to detect mover events and to discriminate between most head orientations, but not between eye orientations (automatic postsurgery gaze-following behavior is predicted to emerge only for head orientation but not for eye orientation); and 3) medium blur conditions, where resolution is sufficient to discriminate between head orientations and also eye orientations, to some degree, in addition to the detection of mover events (the model predicts the emergence of preoperative automatic gaze-following behavior from both head and eye orientation, which may become gradually refined after surgery). These predictions could be evaluated in more detail in future studies using additional data, including preoperative tests of detecting mover events as well as head and eye orientation.

Another open question is whether an active rehabilitation program might enable eye-based gaze understanding and gaze following. Specifically, recreating some of the early learning cues found useful in the model (e.g., explicit training, by watching videos of hand-object interactions, and requiring report of the eyes' position at the time of initial object contact) may facilitate learning. We found that such an explicit procedure for learning of eye gaze direction does not immediately generalize to automatic gaze-following behavior (*SI Appendix, Supplementary Text 3*). It remains to be seen if a longer rehabilitation program might allow this.

To conclude, the behavior seen in the late-treated cataract patients following surgery—an improved discrimination of the eyes' orientation in their orbits and a failure to understand and follow eye gaze direction—may arise because the unsupervised gaze-learning process used in infancy is no longer effective. The failure to acquire a visual behavior or function despite an improvement in visual acuity allowing for better pattern detection may be more general in nature beyond gaze following. It may be of interest to explore in future studies other possible limitations of restored vision caused because unsupervised learning processes guiding early visual learning are no longer available [e.g., precise object boundaries (46) or visual extraction of certain spatial relations such as containment (47)].

Methods

Participants. Fifteen Ethiopian children (age 12.6 ± 3.7 y) with early-onset bilateral cataracts who were surgically treated years later participated in the gaze-cueing experiment. Our sample size was severely limited because of the rareness of the condition (untreated isolated congenital bilateral cataracts), and our inclusion criteria: the ability to recognize others' faces and eyes (as such) and detect a gaze change. The presence of nystagmus and family history (when available) suggests that the cataracts were before 6 mo/congenital. The children underwent cataract removal surgery and insertion of intraocular lens implants in Hawassa Referral Hospital (age at operation of 11.1 ± 3.6 y). All children were examined by an optometrist, and, when needed, optical correction in the form of glasses was given. The children's guardians gave their written consent for the operation and for participating in the behavioral testing. The late-treated participants performed the behavioral tests 1 mo to 4 y after operation (average of 1.6 ± 1.4 y).

Three late-treated participants, treated only recently, performed the behavioral experiments both before and after surgery. Eleven Israeli children (age 10.1 ± 3.3 y) with congenital bilateral cataracts treated soon after birth (0.4 ± 0.2 y) also participated in the study. A group of 44 Israeli and 2 Ethiopian children (age 8.8 ± 2.3 y) with typical sight development served as the control group.

Nine late-treated children participated in the gazed-object following eye-tracking experiment (age 12.3 ± 2.5 y). Five of them also participated in the gaze-cueing behavioral experiment. Four Ethiopian and 27 Israeli children served as a control group (age 12.4 ± 2.2 y).

Nine late-treated children participated in the "faces" and "people in action" eye-tracking experiment (age 12.3 ± 3.3 y). Eight of them also participated in the cueing behavioral experiment. Seven Ethiopian and four Israeli children served as a control group (age: 11.2 ± 1.9 y).

The procedures were approved by the ethics committee of the Hebrew University of Jerusalem and Hawassa University.

Equipment. The behavioral experiment and the CSF test were carried out using a TP500L Asus laptop with a 15.6-in touchscreen. The screen resolution was $1,366 \times 768$ pixels (or $1,920 \times 1,080$ in later examinations). Cataract-treated participants sat at their preferred distance from the screen (20 to 40 cm), maintaining the same distance in all experiments. Control participants sat at a 40-cm distance from the screen. Experiments were programmed using SR research "Experiment Builder" software.

Visual Acuity Measurement. Before and after operation, all cataract-treated participants underwent a basic acuity test performed by an ophthalmologist to reveal whether a patient can perceive light (LP), see a hand in motion, or count fingers at different distances (FC). However, since this test is noisy and qualitative in nature, it is insufficient for characterizing the residual vision of patients.

Therefore, we relied on our CSF experiment to assess visual acuity, presenting horizontal or vertical gratings at different spatial frequencies and contrasts (*SI Appendix, Fig. 1*). The participant reported the grating orientation. The contrast threshold at each spatial frequency was assessed from performance and plotted to generate the CSF (Fig. 2 B and C). An inverse parabola was fitted to the data and the CSF cutoff frequency (defined as the spatial frequency at which the fitted parabola had a sensitivity of 1) was calculated. Due to screen resolution limits, our maximal spatial frequency was 19.2 cpd (or 13.6 cpd for the screen with poorer resolution) at a viewing distance of 40 cm from the screen.

The correlation between the acuity results of the ophthalmologist test (using the ranking level LP = 1, ..., FC at 3 m = 5) and the CSF measurement was $R = 0.51$.

Behavioral Procedures. The behavioral experiments were preceded by preliminary tests designed to exclude patients that did not have the following required capabilities: 1) face/nonface discrimination, 2) eye region recognition (pointing to the eye in the face; *SI Appendix, Fig. 2*, face/eye recognition), and 3) eye and head gaze change detection. Only participants that exceeded 90% correct responses in these tests were included in the later behavioral tests. All participants (including the three preoperated patients tested) exceeded 90% correct responses in the head gaze change detection. In the eye gaze change detection, some participants struggled, and only participants exceeding 75% (9/12) correct answers ($P = 0.02$) continued to the main experiments. The exception to this rule was the three preoperated participants whose scores were between 55% and 67% correct before surgery (and between 83% and 100% after surgery) but still performed both main experiments (head and eye gaze cueing) before surgery. Note that due to their poor ability to detect a change in eye gaze, their lack of compatibility effect in the eye gaze direction experiment is trivial. Ten control participants (age mean of 10.0 ± 1.3) who were tested in a blurred version of these tests reached 100% correct responses in all conditions. Due to extremely limited time resources, in every outreach visit, we typically began by testing the children with better acuity, who were more likely to pass the inclusion criteria. This selection process clearly skewed our patients' distribution of postoperative acuity in this research toward higher acuity. Six of the late-treated patients tested did not meet the inclusion requirement (detecting eye gaze change in at least 75%). Their postoperative acuity was on average 1.9 cpd (range of 0.9 to 2.7), substantially lower than the ones who passed the inclusion criterion ($M = 5.2$; range of 2.1 to 13.7). Typically, children who passed the inclusion criteria (when tested postoperatively) also had better preoperative visual acuity (Fig. 2C) than

in other late cataract-reversal patient studies, designed to reveal the effect of early patterned visual experience on subsequent visual development.

Two experiments were conducted to study the effects of gaze cueing on RT. Each trial started with a frontal face of an actor directly facing the camera (Fig. 3A). In the head direction experiment, the face was shown wearing sunglasses, such that the eyes could not be seen. In the eye direction experiment, the eyes moved without a change in head position. A touch on the nose of the face initiated the gaze cue after a random interval of 16 to 500 ms. In the head direction experiment, using two frames (33.3 ms), the head direction reached its endpoint facing right or left. Then, 300 ms after the change onset, a balloon appeared in either the right or the left side of the screen, and participants were instructed to touch the balloon as fast as possible. Their touch triggered a sound and the initiation of the next trial. The gaze cue was not predictive of the target position. The experiments consisted of blocks, 32 trials in each block, and only participants who were able to complete at least two blocks of each experiment were included. RTs from correct trials only were used for further analysis. Trials in which the RT was less than 200 ms or greater than the mean RT \pm 3 SD were excluded. The average numbers of trials upon which the effect was computed were 78, 81, 67, and 63 for the eye gaze experiment and 79, 86, 86, and 66 for the head gaze experiment in the controls, early-treated, preoperative, and late-treated groups, respectively.

The pixel resolution of the image stimuli for patients was $1,366 \times 768$, where faces span 330 pixels wide. At a distance of ~ 40 cm from the screen, the faces subtend 8.5 visual degrees, which is equivalent to viewing real faces (about 15 cm wide) from a distance of ~ 1 m.

Participants from the control group viewed a blurred version of the image stimuli. The blurred images were created by convolving the original image with a Gaussian filter kernel. In our experiments, we used a Gaussian filter with a cut-off frequency of 1.6 cpd corresponding to spatial SD of 8 pixels (*SI Appendix, Supplementary Text 1*).

Eye-Tracking Procedures. Eye-tracking data acquisition was done with the SR research portable Eye-Link Duo system with a 17-in presentation laptop viewed at a distance of 55 cm using a chin rest. Measurements were taken from the dominant eye at 500 Hz. Calibration in the cataract-treated group was done manually due to their nystagmus (see details below).

Calibrating the eye-tracking system per individual requires that fixation is stable enough on selected points on the screen (typically nine points). Due to the late-treated participants' nystagmus, eye-tracking calibration was not trivial. Our minimal calibration requirement was having successful fixation (for 300 ms) in three points on the screen forming a triangle in the two upper corners and the mid-lower corner of the screen. This was achieved in 12 late-treated participants, who went on to freely view the images shown in the eye-tracking experiment (9 in the gaze-cueing setting, *Results* and Fig. 4 A and B; 9 in the faces and people in action setting, Fig. 4 B and C; and 6 patients did both tests). Forty-two control participants (31 in the gaze-cueing setting; 11 in the faces and people in action setting) viewed a blurred version of the experiment corresponding to a cutoff frequency of 2.5 cpd, the worst blur level experienced by a member of the subgroup of late-treated patients that qualified for the test. Participants were instructed to remember the images for a future memory test.

In the gaze-cueing setting, participants were shown 42 similarly configured images, including an agent directing his gaze at one of two identical objects placed in front of him (Fig. 4A). As in the behavioral experiment, gaze direction was indicated exclusively by eye direction (eye cue condition) or head direction (head cue condition) in which the agent was wearing dark sunglasses. Each trial was initiated by fixation at a point located below the agent's face, followed by a series of three image frames that generate a vivid perception of (apparent) rotation of the eyes or the head of the agent toward one of the objects (i.e., the congruent object). The rotation was initiated by a fixation on the agent's face or 2 s from the appearance of the first frame, whichever came first. The last image frame, depicting the fully rotated head or eyes and the objects, was "frozen" for 5 s. Three areas of interest (AOIs) were defined in advance per image, including the displayed objects (equal-sized ellipses) and agent's face.

We excluded trials in which the subject did not look at any of the objects or failed to look at the agent's face prior to the objects. The logic here was that effective gaze cueing based on head or eye orientation can only be achieved if one is focusing on the face region first. Participants who had less than 20 valid

trials were excluded from the final analysis (three late-treated patients and four controls).

Data analysis. For each participant per trial, we summed the total fixation time spent looking at the predefined AOIs. This resulted in the values of fixation dwell time spent on the gaze-congruent ("target") and incongruent ("competitor") object per trial. Next, we obtained a normalized dwell time value for each trial, defined as the difference between the congruent and incongruent dwell time values, divided by their sum. Finally, for each subject, the averaged normalized dwell time across all trials was calculated.

In the "faces" and "people in action" eye-tracking setting, two different categories of images were shown. The first, faces, showed a frontal view of a person's face (12 images, Faces of Open Source/Peter Adams). The other category, people in action images, showed a person gazing at an object of action (i.e., tool; 12 images from open-source <https://www.freeimages.com/>). Altogether, there were 24 randomly ordered images, each presented for 3 s. A central fixation point was shown between each stimulus presentation. Before running the experiment, interest areas were defined for the people in action images (including the face of the person, the object of gaze, and another object of a similar size) and for the faces images (the eyes, nose, and mouth of the face) for further post hoc analysis (Fig. 4C, interest areas).

Modeling.

Unsupervised learning of gaze direction in blur conditions. *Learning paradigm.* Computationally, learning to follow the gaze of others associates between an image of a face and the gaze direction, defined as the direction in space from the center of the face (specifically at the eyes) to the target location where the face is looking at. We used a self-guiding model (17), which follows the developmental learning process in young infants, to automatically assign pairs of a face image and a corresponding gaze direction. These pairs were then used to train a neural network (9), which received a face image in its input and yielded the gaze-direction vector in space.

The model. Initial-contact events (termed mover events) guide the model as an internal teaching signal to identify and locate likely gaze targets. The model then associates a nearby face with the presumed gaze direction between the center of the eyes' region in the face and target position. In this study we tested the model when the visual input was highly degraded and, in particular, at the low spatial frequency range, as in the case of cataract patients. The learning mechanism in the model requires that the mover teaching signal is reliable and that there is sufficient information of the face to reliably discriminate gaze directions.

The detection reliability of mover events was evaluated under three blur conditions, in terms of precision, defined as the number of truly detected hand-object contacts (true positives, i.e., tp) out of all the detected mover events ($tp + fp$): precision = $tp/(tp + fp)$. Recall was defined as the number of truly detected hand-object contacts (tp) out of all hand-object contacts in the data ($tp + fn$): recall = $tp/(tp + fn)$ (*SI Appendix, Fig. 5*). The blur was simulated by applying a Gaussian filter to each video frame at the input of the model. We used spatial SD values of 8, 16, and 32 pixels, corresponding to cutoff frequencies of 1.7, 0.8, and 0.4 cpd, respectively. We also used the input at normal resolution with no blur (corresponding to a cutoff frequency of 15.7 cpd) as a control. The data for these simulations included four video sequences (roughly 15 min long; 22,545 frames; 360×288 pixels). The videos show humans sitting behind a table and manipulating objects with their hands and include 135 true mover (hand-object contact) events. The scenes also show autonomously moving objects on the table (e.g., rolling balls).

Prediction of gaze direction under blur conditions. We tested computationally if sufficient information can be extracted from the face image to reliably discriminate gaze directions when the visual input is highly degraded. For this purpose, we trained and evaluated a convolutional neural network in inferring the gaze direction from a given face image. The data consisted of face images extracted from eight video sequences showing eight human actors manipulating objects with their hands while sitting behind a table (roughly 23 min long; 34,631 frames; 540×432 pixels). At normal resolution, a total of 1,549 face images (at $[44 \pm 4] \times [46 \pm 3]$ mean pixel resolution, corresponding to a cutoff frequency of 15.7 cpd) were extracted from the video sequences whenever a mover event was detected (i.e., during initial pick-up or put-down interaction), with the assumption that the actors look at objects they manipulate. Each face image was assigned with the gaze direction between the image location of the

center of the eye region in the face and the location of the target hand-object contact event. In our simulations, the gaze direction was represented as the signed angle between the vertical y axis and the vector connecting the face to the target location.

We used a deep network based on Resnet-50 (10) architecture trained to identify faces on the VGG-Face2 (48) dataset. We first retrained the network to identify the eight identities of the actors in our dataset (5 epochs, 20 cropped augmentations per train image, 25 test images per identity) and then utilized a transfer learning procedure to train the network to predict a gaze-direction vector from a given face image input. Simulations consisted of a training phase with half of our dataset (all face images of 4 face identities, 50 epochs, 20 cropped augmentations per image) and an evaluation phase on the remaining face images of the other 4 face identities (demonstrating the full generalization of the network across face identities).

Discrimination of head and eye orientation from face representations in restored acuity conditions. Evaluation paradigm. We tested if face representations developed with high blur input, similar to preoperative conditions, are sufficient to automatically discriminate head and eye orientations (without explicit learning) given input at restored acuity, similar to postoperative conditions. In particular, we used a deep convolutional neural network trained to recognize face identities in images. The network's architecture is based on ResNet (10), which was shown to create rich image feature maps and proved to be very useful in visual tasks, including object classification and face identification. For the evaluation, face representations were extracted from the network's internal activation levels and were used to discriminate extreme (left versus right) head and eye orientations (Fig. 6A).

Learning face representations. In our simulations, the neural network was trained from scratch (i.e., with initial random weights at all layers) to recognize face identities in natural images (but not face or eye orientations) using the VGGFace2 (48) dataset. Training was done in the high-blur regimen, corresponding to a cutoff frequency of 0.8 cpd (5 epochs; 3.31 M images; 9,131 face identities; average of 363 images per face identity; mean face identification accuracy of 55%). This high-blur regimen simulates the blur conditions experienced by cataract patients before and immediately after treatment. Similar to patients with prolonged visual deprivation, the network was not exposed to high frequencies in the input.

Discriminating head and eye orientations under high-blur conditions. Face images were fed to the input of the neural network, and face representations were extracted from the network as the activation levels of the second intermediate layer ("layer 2"; output vector size is 512). With no further training, we tested if the extracted face representations encode sufficient information for discriminating between faces looking to the left and to the right. We used two sets of images (336 images in a set) with both faces looking to the left and to the right. One set of images showed faces oriented to the left or to the right. The other set showed frontal faces with eyes shifted left or right in their orbits. Both sets were extracted from the Columbia Gaze dataset (49). For each image set, we measured mean classification accuracy of a support vector machine binary classifier (left/right) applied to the face representation vectors using a leave-one-

out cross-validation technique. All images were blurred at the input of the network at the same conditions used in the training.

Discriminating head and eye orientations under restored acuity conditions. To simulate the patients' poor visual experience before surgery and improved acuity, shortly after surgery, we repeated the evaluation process described in the previous section above. This time, however, we applied the previously trained neural network (on low-pass-filtered images at 0.8 cpd) with no further training, to images containing higher spatial frequencies (i.e., images at finer resolution). These images were filtered with a Gaussian kernel (spatial SD of 4 pixels), corresponding to a cutoff frequency of 3.3 cpd, similar to the worst acuity of postsurgery patients.

Further learning of face representations under restored acuity conditions. Lastly, we simulated the patients' visual experience long after surgery, when their visual system could adapt to the improved acuity at the input. In this simulation, the network that was trained with highly blurred input was further trained with higher frequencies in the input (i.e., with input images at higher resolution [five epochs; mean face identification accuracy of 84%]). In our experiment, we used a Gaussian filter with an SD of 4 pixels (corresponding to a cutoff frequency of 3.3 cpd). The new face representations with further adaptation to the higher frequencies in the input were then used to discriminate head and eye orientations (as before with no explicit training) in input images at the same level of acuity (i.e., with a cutoff frequency of 3.3 cpd).

Control condition (no blur). For comparison, we performed the above evaluation procedure (learning face representations and using them to discriminate head and eye orientations) with a control neural network that was trained from scratch with input images at normal resolution and with no blur (corresponding to a cutoff frequency of 17.6 cpd), similar to the visual experience of control subjects with normal vision (five epochs; mean face identification accuracy of 75%).

Data Availability. Data are available in the *SI Appendix* and Mendeley Data (DOI: [10.17632/58p59bkjfv.1](https://doi.org/10.17632/58p59bkjfv.1)) (30).

ACKNOWLEDGMENTS. We thank our participants and Ms. Zemene Zeleke for her invaluable help in organizing the medical treatment and testing of the children in Ethiopia. We thank Lior Aloni, Daniel Houry, and Noam Behar for their help in data collection. The Jerusalem science museum generously provided us the facility to test normally sight-developing children. This study was supported by the DFG German-Israel cooperation grant Z0 349/1 to E.Z. and S.U. and by a research grant from the Carolito Stiftung and the Robin Neustein Artificial Intelligence research fellowship.

Author affiliations: ^aThe Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; ^bDepartment of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel; ^cDepartment of Ophthalmology, Padeh Medical Center, Poriya 15208, Israel; ^dDepartment of Optometry and Vision Science, Hadassah Academic College, Jerusalem 91010, Israel; and ^eNeurology Department, Hadassah Medical Organization and Faculty of Medicine, Jerusalem 91120, Israel

1. R. Brooks, A. N. Meltzoff, "Gaze following: A mechanism for building social connections between infants and adults" in *Mechanisms of Social Connection: From Brain to Group*, M. Mikulincer, P. R. Shaver, Eds. (American Psychological Association, 2013), pp. 167-183.
2. C. Moore, V. Corkum, Social understanding at the end of the first year of life. *Dev. Rev.* **14**, 349-372 (1994).
3. R. Brooks, A. N. Meltzoff, The importance of eyes: How infants interpret adult looking behavior. *Dev. Psychol.* **38**, 958-966 (2002).
4. M. Suzuki, A. Izawa, K. Takahashi, Y. Yamazaki, The coordination of eye, head, and arm movements during rapid gaze orienting and arm pointing. *Exp. Brain Res.* **184**, 579-585 (2008).
5. M. Tomasello, B. Hare, H. Lehmann, J. Call, Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *J. Hum. Evol.* **52**, 314-320 (2007).
6. S. Spadacenta, P. W. Dicke, P. Thier, Reflexive gaze following in common marmoset monkeys. *Sci. Rep.* **9**, 15292 (2019).
7. S. V. Shepherd, R. O. Deaner, M. L. Platt, Social status gates social attention in monkeys. *Curr. Biol.* **16**, R119-R120 (2006).
8. J. Call, M. Tomasello, Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* **12**, 187-192 (2008).
9. A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet classification with deep convolutional neural networks" in *Proceedings of the 25th International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2012), pp. 1097-1105.
10. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, 2016), pp. 770-778.
11. K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask R-CNN" in *Proc. IEEE Int. Conf. Comput. Vis.* (IEEE, 2017), pp. 2980-2988.
12. Q. V. Le et al., "Building high-level features using large scale unsupervised learning" in *Proceedings of the 29th International Conference on Machine Learning* (Omnipress, 2012), pp. 81-88.
13. A. R. Contintente, A. Khosla, C. Vondrick, Where are they looking? *Adv. Neural Inf. Process. Syst.* **28**, 199-207 (2015).
14. E. Wood et al., "Rendering of eyes for eye-shape registration and gaze estimation" in *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2015), pp. 3756-3764.
15. A. Shrivastava et al., "Appearance-based gaze estimation in the wild" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 4511-4520.
16. E. Murphy-Chutorian, M. M. Trivedi, Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 607-626 (2009).
17. S. Ullman, D. Harari, N. Dorfman, From simple innate biases to complex visual concepts. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18215-18220 (2012).
18. R. Saxe, J. B. Tenenbaum, S. Carey, Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychol. Sci.* **16**, 995-1001 (2005).
19. S. Amano, E. Kezuka, A. Yamamoto, Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behav. Dev.* **27**, 64-80 (2004).
20. V. Slaughter, P. Neary, Do young infants respond socially to human hands? *Infant Behav. Dev.* **34**, 374-377 (2011).
21. A. Kalia et al., Development of pattern vision following early and extended blindness. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2035-2039 (2014).

22. T. Farroni, G. Csibra, F. Simion, M. H. Johnson, Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9602-9605 (2002).
23. M. D. Vida, D. Maurer, The development of fine-grained sensitivity to eye contact after 6 years of age. *J. Exp. Child Psychol.* **112**, 243-256 (2012).
24. B. M. Hood, J. Douglas Willen, J. Driver, Adult's eyes trigger shifts of visual attention in human infants. *Psychol. Sci.* **9**, 131-134 (1998).
25. T. Farroni, S. Massaccesi, D. Pividori, M. H. Johnson, Gaze following in newborns. *Infancy* **5**, 39-60 (2004).
26. M. Scaife, J. S. Bruner, The capacity for joint visual attention in the infant. *Nature* **253**, 265-266 (1975).
27. D. Maurer, P. Salapatek, Developmental changes in the scanning of faces by young infants. *Child Dev.* **47**, 523-527 (1976).
28. B. Röder, R. Kekunnaya, Visual experience dependent plasticity in humans. *Curr. Opin. Neurobiol.* **67**, 155-162 (2021).
29. C. K. Friesen, A. Kingstone, The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychol. Sci.* **26**, 393-401 (2015).
30. E. Zohary *et al.*, Gaze-following requires early visual experience data. Mendeley Data. 10.17632/58p59bkjfv.1. Deposited 11 April 2022.
31. S. Ganesh *et al.*, Results of late surgical intervention in children with early-onset bilateral cataracts. *Br. J. Ophthalmol.* **98**, 1424-1428 (2014).
32. J. Ye *et al.*, Resilience of temporal processing to early and extended visual deprivation. *Vision Res.* **186**, 80-86 (2021).
33. F. W. Campbell, J. G. Robson, Application of Fourier analysis to the visibility of gratings. *J. Physiol.* **197**, 551-566 (1968).
34. D. Maurer, C. J. Mondloch, T. L. Lewis, Effects of early visual deprivation on perceptual and cognitive development. *Prog. Brain Res.* **164**, 87-104 (2007).
35. N. K. Pehera, G. Chandrasekhar, R. Kekunnaya, The critical period for surgical treatment of dense congenital bilateral cataracts. *J. AAPOS* **13**, 527-528 (2009).
36. P. Zerr *et al.*, Successful visually guided eye movements following sight restoration after congenital cataracts. *J. Vis.* **20**, 3 (2020).
37. D. Harari, T. Gao, N. Kanwisher, J. Tenenbaum, S. Ullman, Measuring and modeling the perception of natural and unconstrained gaze in humans and machines. arXiv [Preprint] (2016). <https://doi.org/10.48550/arXiv.1611.09819> (Accessed 6 April 2022).
38. S. Sourav, D. Bottari, I. Shareef, R. Kekunnaya, B. Röder, An electrophysiological biomarker for the classification of cataract-reversal patients: A case-control study. *EClinicalMedicine* **27**, 100559 (2020).
39. A. Senju *et al.*, Early social experience affects the development of eye gaze processing. *Curr. Biol.* **25**, 3086-3091 (2015).
40. R. Le Grand, C. J. Mondloch, D. Maurer, H. P. Brent, Impairment in holistic face processing following early visual deprivation. *Psychol. Sci.* **15**, 762-768 (2004).
41. C. L. Grady, C. J. Mondloch, T. L. Lewis, D. Maurer, Early visual deprivation from congenital cataracts disrupts activity and functional connectivity in the face network. *Neuropsychologia* **57**, 122-139 (2014).
42. A. McKyton, I. Ben-Zion, E. Zohary, Lack of automatic imitation in newly sighted individuals. *Psychol. Sci.* **29**, 304-310 (2018).
43. Y. Ostrovsky, A. Andalman, P. Sinha, Vision following extended congenital blindness. *Psychol. Sci.* **17**, 1009-1014 (2006).
44. R. Flom, K. Lee, D. Muir, *Gaze-Following: Its Development and Significance* (Psychology Press, 2017).
45. B. D'Entremont, S. M. J. Hains, D. W. Muir, A demonstration of gaze following in 3- to 6-month-olds. *Infant Behav. Dev.* **20**, 569-572 (1997).
46. N. Dorfman, D. Harari, S. Ullman, "Learning to perceive coherent objects" in *Proceedings of the Annual Meeting of the Cognitive Science Society*, M. Knauff, M. Pauen, N. Sebanz, I. Wachsmuth, Eds. (Cognitive Science Society, Austin, TX, 2013), pp. 394-399.
47. S. Ullman, N. Dorfman, D. Harari, A model for discovering 'containment' relations. *Cognition* **183**, 67-81 (2019).
48. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age" in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (IEEE, 2018), pp. 67-74.
49. B. A. Smith, Q. Yin, S. K. Feiner, S. K. Nayar, "Gaze locking: Passive eye contact detection for human-object interaction" in *Proceedings of the 26th annual ACM Symposium on User Interface Software and Technology* (ACM, October 2013), pp. 271-280.