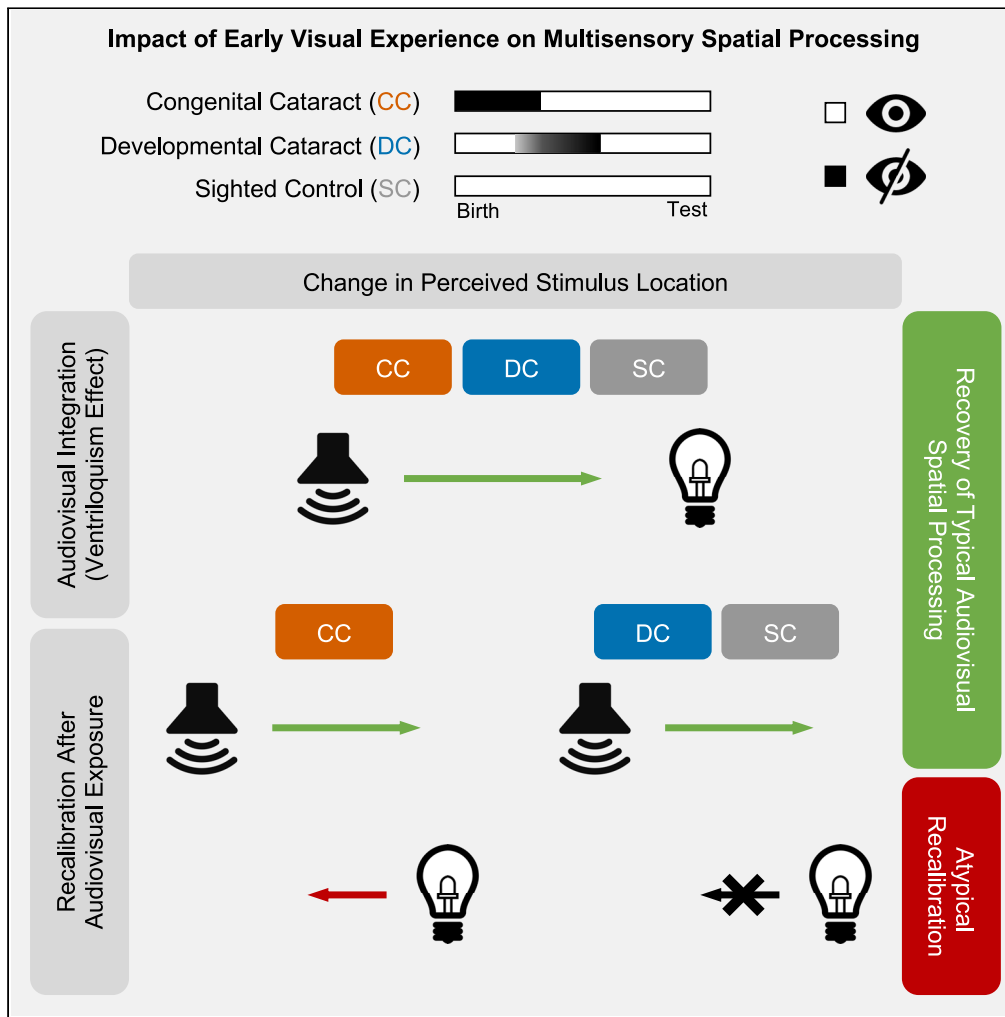


Article

Audiovisual spatial recalibration but not integration is shaped by early sensory experience



Patrick Bruns, Lux Li, Maria J.S. Guerreiro, ..., Kabilan Pitchaimuthu, Ramesh Kekunnaya, Brigitte Röder

patrick.bruns@uni-hamburg.de

Highlights
 Audiovisual spatial processing recovers after congenital cataract removal surgery

Long durations of congenital blindness do not prevent recovery

Sight-recovery individuals use audiovisual input to recalibrate visual localization

Typical and atypical crossmodal mechanisms coexist after sight recovery

Bruns et al., iScience 25, 104439
 June 17, 2022 © 2022 The Author(s).
<https://doi.org/10.1016/j.isci.2022.104439>



Article

Audiovisual spatial recalibration but not integration is shaped by early sensory experience

Patrick Bruns,^{1,5,*} Lux Li,^{1,2} Maria J.S. Guerreiro,^{1,3} Idris Shareef,⁴ Siddhart S. Rajendran,^{1,4} Kabilan Pitchaimuthu,^{1,4} Ramesh Kekunnaya,⁴ and Brigitte Röder¹

SUMMARY

To clarify the role of sensory experience during early development for adult multisensory learning capabilities, we probed audiovisual spatial processing in human individuals who had been born blind because of dense congenital cataracts (CCs) and who subsequently had received cataract removal surgery, some not before adolescence or adulthood. Their ability to integrate audio-visual input and to recalibrate multisensory spatial representations was compared to normally sighted control participants and individuals with a history of developmental (later onset) cataracts. Results in CC individuals revealed both normal multisensory integration in audiovisual trials (ventriloquism effect) and normal recalibration of unimodal auditory localization following audiovisual discrepant exposure (ventriloquism aftereffect) as observed in the control groups. In addition, only the CC group recalibrated unimodal visual localization after audiovisual exposure. Thus, in parallel to typical multisensory integration and learning, atypical cross-modal mechanisms coexisted in CC individuals, suggesting that multisensory recalibration capabilities are defined during a sensitive period in development.

INTRODUCTION

Sensory experiences made during sensitive periods of development have a strong and lasting impact on how information from the different sensory systems is combined (Putzar et al., 2007; Schorr et al., 2005; Smyre et al., 2021; Sourav et al., 2019; Wallace and Stein, 2007). For example, neuroanatomical and neurophysiological studies in owls have revealed distinct neural mechanisms for crossmodal recalibration depending on the age of the animals: Although crossmodal recalibration in juvenile owls was mediated by anatomical changes (DeBello et al., 2001; Feldman and Knudsen, 1997), it was associated with top-down guided physiological changes in adult animals (Hyde and Knudsen, 2002; Knudsen, 1998).

These different neural mechanisms of crossmodal learning in the developing and adult system were impressively demonstrated in an experiment in which juvenile owls were exposed to a short phase of prism experience, resulting in a shift of perceived auditory locations to correct for the visual spatial disparity, and consecutively grew up with typical visual input. When the prisms were refitted in adulthood these animals demonstrated, unlike adult owls without any prism experience early in life, juvenile-like recalibration of auditory spatial maps. Crucially, this prevailing learning capacity was limited to the same prism-mediated shift of the visual world that was experienced during development (Knudsen, 1998; Linkenhoker et al., 2005). These results, hence, suggest that the involved neural systems were not more plastic in general, but that these animals had access to specific additional anatomical connections which had been established during the juvenile prism experience. Thus, it seems that anatomical connectivity acquired during the sensitive period sets the limits of functional adaptation in adulthood and that the maintenance of these connections is at least partially independent of use following the sensitive period.

Behavioral findings in humans with late-onset blindness provide support for this hypothesis as well. Despite dramatic changes in the accessible sensory world, late blind humans have been shown to use the typical visually defined spatial reference frame for representing touch in tasks in which congenitally blind humans use somatotopically defined (Collignon et al., 2009; Röder et al., 2004) or head-centered references frames (Röder et al., 2007), even if blindness had lasted for more years than in the congenitally blind participants of the same study. Similarly, it has been reported that congenitally blind humans lack typical sound-shape

¹Biological Psychology and Neuropsychology, University of Hamburg, 20146 Hamburg, Germany

²Department of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, ON N6G 2M1, Canada

³Biological Psychology, Department of Psychology, School of Medicine and Health Sciences, University of Oldenburg, 26111 Oldenburg, Germany

⁴Jasti V Ramanamma Children's Eye Care Centre, LV Prasad Eye Institute, Hyderabad, Telangana 500034, India

⁵Lead contact

*Correspondence: patrick.bruns@uni-hamburg.de
<https://doi.org/10.1016/j.isci.2022.104439>



associations observed in sighted participants (Fryer et al., 2014; Hamilton-Fletcher et al., 2018; Sourav et al., 2019), whereas late blind humans still demonstrate them many years after total blindness had started (Sourav et al., 2019).

If crossmodal connectivity is indeed acquired during an early sensitive phase of development and stabilized in a way that prevents loss later in life, individuals who were born blind but recovered sight later in life should demonstrate altered multisensory processes reminiscent of those observed in congenitally blind humans. Moreover, multisensory interactions of the newly acquired visual system with the auditory and tactile systems would be expected to differ. The evidence currently available draws an inconsistent picture: Individuals who had been born blind because of bilateral dense cataracts and who had undergone cataract removal surgery later in life demonstrated processing gains from redundant simple crossmodal stimulation (De Heering et al., 2016; Putzar et al., 2007, 2012) and even showed visual-haptic integration as assessed by the precision of size estimates (Senna et al., 2021) or the occurrence of the size-weight illusion (Pant et al., 2021). However, they were less able to integrate audio-visual speech input (Putzar et al., 2007, 2010), did not show typical crossmodal sound-shape associations (Sourav et al., 2019), and used spatial reference frames typical for congenitally blind humans (Azanon et al., 2018; Ley et al., 2013; cf. Collignon et al., 2009; Röder et al., 2004).

Altered multisensory processing might be related to impaired multisensory learning mechanisms such as crossmodal recalibration. Recent findings have suggested two dissociable types of crossmodal recalibration which emerge at two distinct time scales: (1) Immediate recalibration on a trial-by-trial basis following a single exposure to a discrepant crossmodal stimulus and (2) cumulative recalibration after prolonged exposure to a consistent crossmodal discrepancy (Bruns and Röder, 2015; Park and Kayser, 2021; Watson et al., 2019). However, whether or how immediate and cumulative multisensory learning mechanisms depend on early sensory input has not been investigated in humans yet. A study in children has recently found that crossmodal spatial recalibration does not arise before middle childhood (with immediate recalibration occurring before cumulative recalibration) and, thus, develops later than the ability to integrate auditory and visual spatial input for localization (Rohlf et al., 2020). Moreover, it has been suggested that the weights assigned to each sensory cue during cumulative recalibration in adulthood do not depend on their precision but rather seem to be fixed (Rohlf et al., 2021; Zaidel et al., 2011; but see Hong et al., 2021), whereas they are dependent on cue reliability in both immediate recalibration and multisensory integration (Rohlf et al., 2021). Thus, the late developmental onset of cumulative crossmodal recalibration might be a consequence of the need to acquire the relative weighting of individual sensory cues through extensive experience which is consecutively stabilized by an elaboration of anatomical connectivity during development (Linkenhoker et al., 2005).

Multisensory integration requires not only a proper weighting of the sensory cues but also a decision of whether the sensory cues arise from a common cause and, thus, should be integrated at all (Kayser and Shams, 2015). Humans seem to solve this causal inference problem implicitly and optimally (or near-optimally) as predicted by Bayesian hierarchical causal inference (CI) models (Körding et al., 2007; Wozny et al., 2010), which account for both the sensory cue binding and weighting aspects by assuming that two perceptual estimates are derived, one under the assumption that the sensory cues originated from a single source (fusion) and one under the assumption that they had separate sources (segregation). Previous studies have suggested some variability across individuals in the decision strategy that is used to integrate these two estimates (Wozny et al., 2010): The majority of participants selected one of the two causal structures in proportion to their probability (probability matching), but in some participants the behavioral responses were better described by a weighted averaging of the two estimates (model averaging) or by the selection of the most likely causal structure (model selection). Regardless of the specific individual decision strategy used, the superiority of the CI models in explaining human behavior, as compared to a model that only takes cue reliability but not the causal structure into account (forced fusion models), suggests that multisensory integration typically incorporates both causal inference and cue reliability.

If early crossmodal experience is critical for multisensory spatial integration and recalibration capabilities to emerge, we would expect altered multisensory spatial processing in CC reversal individuals because of the lack of visual (and as a consequence altered crossmodal) input after birth. The present study tested this hypothesis by investigating both multisensory integration and crossmodal recalibration in individuals who had been born blind because of bilateral dense cataracts which were removed between the ages of

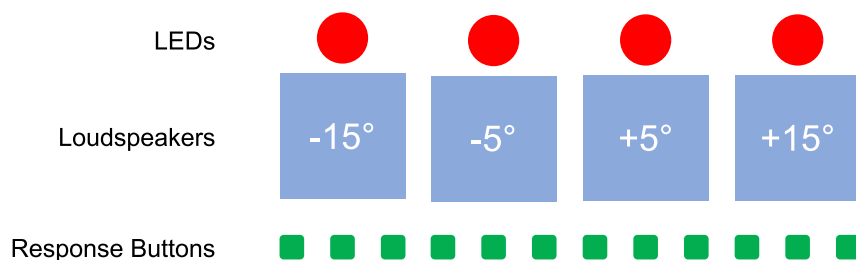


Figure 1. Schematic illustration of the experimental setup

Participants faced the center of four loudspeakers positioned at $\pm 15^\circ$ and $\pm 5^\circ$. Red LEDs were attached on top of each loudspeaker. An array of 12 pushbuttons, which had the same width as the loudspeaker array, was placed in front of the loudspeakers.

5 months and 33 years. Although congenital bilateral dense cataracts typically do not result in total blindness (i.e., a loss of light perception), the CC reversal individuals selected for the present study fulfilled the criteria of blindness according to the classification of the World Health Organization (2019) before cataract-removal surgery (i.e., visual acuity worse than 3/60).

We used a typical and well-established ventriloquism paradigm with simple sounds and visual stimuli which were presented with a spatial disparity. The participants' task was to localize both the auditory and the visual component of the crossmodal stimuli, which allowed us to assess the ability to spatially integrate audiovisual input after reversal of a congenital blindness. Additional unimodal visual and auditory test trials were included to assess the immediate (VAEi) and cumulative ventriloquism aftereffects (VAEc) indicating crossmodal recalibration (for a schematic illustration of the experimental setup, see Figure 1). We expected a smaller ventriloquism effect (VE), and relatedly a smaller VAEi, because of prevailing visual deficits in CC reversal individuals compared to normally sighted controls. If normative causal inference and optimal multisensory integration require exposure to audiovisual correspondences during a sensitive phase of early development, we would also expect the localization performance of CC reversal individuals to be less well described by a CI model. Moreover, if the relative weighting of individual cues for cumulative crossmodal recalibration is acquired during a sensitive phase of development, we would expect a lower influence of vision on the recalibration of auditory spatial representations but a higher influence of hearing on recalibration of visual spatial representations in CC reversal individuals compared to normally sighted controls and individuals who had been treated for developmental (i.e., late onset) cataracts.

RESULTS

To determine the influence of early visual experience on multisensory integration (as indicated by the VE) and learning (as indicated by the VAE) in later life, we compared a group of 11 individuals who had been born blind because of dense bilateral congenital cataracts (CC group), and who underwent cataract-removal surgeries later in life (between 5 months and 33 years of age), with two control groups, one group of 10 normally sighted control participants (SC group) and one group of 10 individuals with developmental cataracts (DC group). All participants had to localize auditory (A) and visual (V) stimuli coming from four different azimuthal positions. A and V stimuli were either presented together with different audiovisual (AV) spatial discrepancies (adaptation trials used to assess the VE) or were presented alone as unimodal test trials to assess immediate recalibration effects (VAEi), that is, a shift in sound or visual localization after a single exposure to a spatially discrepant AV stimulus. In a second block, the AV spatial discrepancy was fixed at 10° to assess cumulative recalibration effects (VAEc) in the unimodal test trials.

Early visual input is not necessary for the development of audiovisual spatial integration (VE)

In AV trials, participants had to localize both the A and the V stimulus components as in previous studies of the VE (Jackson, 1953; Körding et al., 2007; Mohl et al., 2020; Wozny et al., 2010; Wozny and Shams, 2011). The degree to which these responses were modulated by the AV spatial discrepancy (i.e., the size of the VE) is, thus, an indicator of multisensory integration. Here we defined the VE as the difference in localization responses between AV trials with a rightward discrepancy (V to the right from location A) and AV trials with a leftward discrepancy (V to the left from location A). The resulting VE values are shown in Figure 2 (for group-averaged responses as a function of AV discrepancy, see also Figure S1).

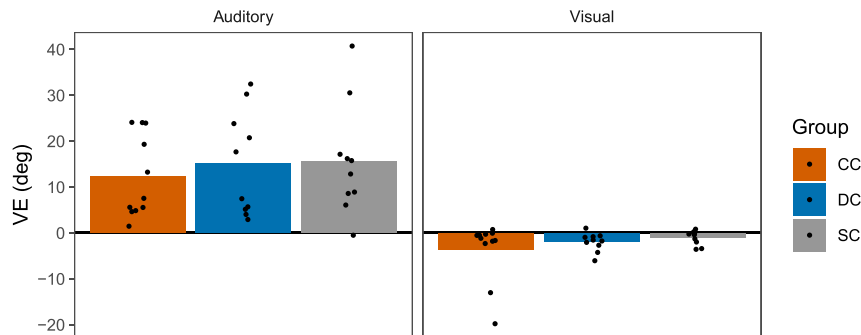


Figure 2. Ventriloquism effects (VE) in congenital cataract (CC), developmental cataract (DC), and sighted control (SC) individuals

The differences in localization responses (in degrees) between audiovisual (AV) trials with a rightward spatial discrepancy (V to the right of A) and AV trials with a leftward discrepancy (V to the left of A) are shown separately for each group and response modality. Bars indicate group averages and dots indicate individual values. Positive values indicate localization biases toward the V location and negative values indicate localization biases toward the A location. (See also Figures S1–S4, S7, and S8).

In all three groups, A responses were clearly biased toward the V location [$p \leq 0.002$; $BF_{+0} \geq 36.21$], indicative of an auditory VE. The size of the auditory VE did not significantly differ between groups [$F(2,17.90) = 0.33$, $p = 0.720$; $BF_{10} = 0.25$]. Conversely, V responses tended to be biased toward the A location. As for the auditory VE, there was no significant difference in the size of the visual VE between groups [$F(2,17.11) = 1.39$, $p = 0.276$; $BF_{10} = 0.43$], although the visual VE only reached significance in the DC group [$p = 0.019$, $BF_{-0} = 10.04$] and not in the CC and SC groups [$p \geq 0.080$, $BF_{-0} \leq 2.35$]. Thus, all three groups exhibited similar multisensory integration capabilities in the AV trials of Block 1.

To further characterize multisensory integration in all groups, we next modeled the spatial distributions of A and V responses as a function of the AV discrepancy (see Figures S2–S5). We fitted and compared three Bayesian hierarchical causal inference (CI) models of multisensory perception, each with a different decision strategy (Wozny et al., 2010): model averaging (CI-MA), probability matching (CI-PM), and model selection (CI-MS), and four non-CI baseline models to each participant's Block 1 AV localization responses (ventriloquism effect). The CI models accounted for uncertainty in the causal structure that could give rise to the A and V cues, that is, the relative probability of the same versus different sources causing the cues was assessed to determine whether and how to integrate the cues (Körding et al., 2007). Unlike the CI models, the baseline models assumed a fixed causal structure and either always integrated (forced fusion, FF) or always segregated (SG) the A and V cues to estimate the stimulus locations. The FF and SG models each had two forms: either incorporating a Gaussian location prior (FF, SG) or no priors (FFnp, SGnp). The best model for an individual participant was selected based on having the lowest Bayesian information criterion (BIC) value (Tables 1 and S1). No participant was best fitted by model FFnp or SGnp. A Chi-square test of independence on the numbers of participants best fitted by the other five models indicated that the distribution of best-fitting models did not depend on the group [$\chi^2(8) = 11.42$, $p = 0.179$].

Overall, model CI-PM predicted the AV localization responses best, as it was the best-fitting model for 17 of 31 (54.8%) participants in the total sample including the majority of CC and SC participants (Table S1), and had the lowest overall BICs for CC and SC participants (Table 1). CI-MS accounted for most of the remaining participants (22.6%) and was the best-fitting model in DC participants, with CI-PM performing only slightly worse than CI-MS in this group (Table 1). Overall, this finding is similar to previous behavioral studies of the VE in normally sighted human participants, which found that CI-PM explained the AV spatial localization data better than CI-MA and CI-MS (Wozny et al., 2010; Wozny and Shams, 2011). Importantly, in all three groups the BIC values indicated that all three CI models described the data substantially better than each of the four non-CI models. The forced fusion model with no priors (FFnp), commonly known as the optimal integration model (Alais and Burr, 2004; Ernst and Banks, 2002), described the data worst and did not improve much even after incorporating the location prior (Table 1). In summary, the observed AV localization pattern of most participants, including the majority of CC participants, was highly consistent with Bayesian causal inference, as evident by the overall superior performance of the CI models compared

Table 1. Comparison of model performances

	CI-MA	CI-PM	CI-MS	FF	SG	FFnp	SGnp
CC (n = 11)	836.97 ± 118.11 (3)	797.99 ± 114.67 (1)	813.68 ± 115.86 (2)	1761.80 ± 147.40 (6)	961.10 ± 139.48 (4)	1681.48 ± 147.35 (7)	985.67 ± 137.35 (5)
DC (n = 10)	1010.54 ± 83.77 (3)	971.98 ± 77.51 (2)	967.90 ± 78.76 (1)	1808.48 ± 166.79 (6)	1051.49 ± 97.44 (4)	1814.98 ± 167.54 (7)	1061.66 ± 97.42 (5)
SC (n = 10)	913.16 ± 116.04 (3)	880.39 ± 110.71 (1)	892.67 ± 107.29 (2)	1859.86 ± 266.33 (6)	1031.62 ± 120.40 (4)	1875.26 ± 262.58 (7)	1059.65 ± 116.43 (5)

The table shows the mean ± SEM Bayesian information criterion (BIC) values of model fits to Block 1 audiovisual localization data across individually fitted participants. A lower BIC indicates better model performance for a given dataset. Bracketed numbers indicate each model's ranking within a group. Boldfaced is the overall best model within each group. According to Raftery (1995), a BIC difference of 0–2 between two models suggests weak evidence as to which model is better, a BIC difference of 2–6 shows positive evidence (in favor of the model with the lower BIC), a BIC difference of 6–10 shows strong evidence, and a BIC difference >10 shows very strong evidence. Being compared are three causal inference models (CI-MA: causal inference with model averaging; CI-PM: causal inference with probability matching; CI-MS: causal inference with model selection) and four baseline models (FF: forced fusion with location prior; SG: segregation with location prior; FFnp: forced fusion with no priors; SGnp: segregation with no priors). (See also Tables S1 and S2 and Figures S2–S5).

with the baseline models. These findings suggest that multisensory integration in CC individuals was guided by similar computational principles as in the control groups.

Immediate crossmodal recalibration (VAEi) recovers after restoring sight in congenital blindness

To assess whether participants immediately recalibrated their unimodal localization in response to spatially discrepant AV stimuli on a trial-by-trial basis, we computed the difference in localization responses between unimodal trials (A and V trials, respectively) which were preceded by an AV trial with a rightward discrepancy and unimodal trials which were preceded by an AV trial with a leftward discrepancy. The resulting VAEi values are shown in Figure 3 (for group-averaged responses as a function of the spatial discrepancy in the preceding AV trial, see also Figure S6).

In all three groups, localization responses in A trials were biased in the direction of the preceding AV discrepancy [$p \leq 0.018$, $BF_{+0} \geq 5.68$], indicating a typical auditory VAEi. The size of the auditory VAEi did not significantly differ between groups [$F(2, 18.34) = 0.32$, $p = 0.728$; $BF_{10} = 0.25$]. In contrast to the auditory VAEi, localization responses in V trials were not significantly modulated by the preceding AV spatial discrepancy in any of the three groups [$p > 0.999$, $BF_{-0} \leq 0.41$] and there were no significant group differences [$F(2, 17.35) = 2.10$, $p = 0.153$; $BF_{10} = 0.54$]. Thus, on a trial-by-trial basis, all three groups recalibrated auditory but not visual localization in response to immediately preceding spatially discrepant AV stimuli.

Across groups, the size of the auditory VAEi was significantly correlated with the size of the auditory VE [$r_s = 0.49$, $p = 0.005$], suggesting a link between multisensory integration and immediate crossmodal recalibration (see Figure 4). The correlation between auditory VE and VAEi was significant within the group of CC individuals [$p = 0.012$], but not within the DC and SC groups [$p \geq 0.494$].

Early sensory experience defines the role of vision and hearing for calibrating multisensory spatial representations (VAEc)

To assess whether participants recalibrated their unimodal localization in response to cumulative evidence for a crossmodal spatial mismatch, we calculated the difference in unimodal localization responses (A and V trials, respectively) between Block 2, in which AV trials featured a constant spatial discrepancy of 10°, and unimodal localization responses in Block 1, in which AV trials had a mean spatial discrepancy of 0°. The resulting VAEc values (see Figure 5) correspond to changes in the likelihood mean bias parameters (Δ_A , Δ_V) of our computational models from Block 1 to Block 2 (for tests of changes in other model parameters, see Table S2 and Figures S2–S4).

In the CC and SC groups, localization responses in A trials were significantly biased in the direction of the constant AV spatial discrepancy in Block 2 [$p \leq 0.006$, $BF_{+0} \geq 24.30$], indicating a typical auditory VAEc. The

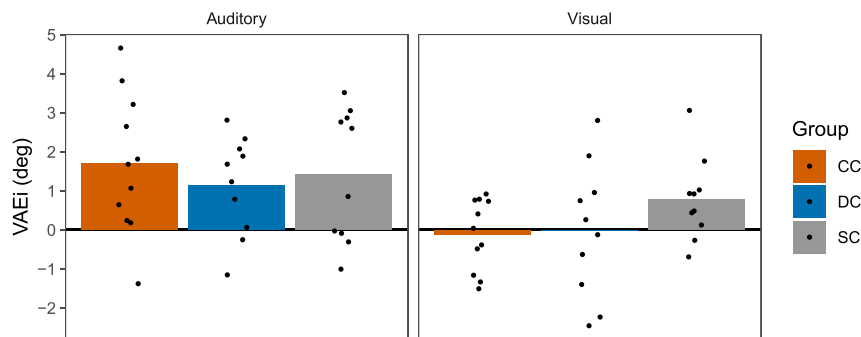


Figure 3. Immediate ventriloquism aftereffects (VAEi) in CC, DC, and SC individuals

The differences in localization responses (in degrees) between unimodal trials (A and V trials, respectively) which were preceded by an AV trial with a rightward discrepancy (V to the right of A) and unimodal trials which were preceded by an AV trial with a leftward discrepancy (V to the left of A) are shown separately for each group and modality. Bars indicate group averages and dots indicate individual values. Positive values indicate localization biases toward the V location and negative values indicate localization biases toward the A location. (See also Figures S6–S8).

auditory VAEc was not significant in the DC group [$p = 0.111$, $BF_{+0} = 1.06$]. Despite this, the size of the auditory VAEc did not significantly differ between groups [$F(2, 17.84) = 0.65$, $p = 0.532$; $BF_{10} = 0.36$]. CC individuals also showed a highly significant visual VAEc in V trials [$p = 0.003$; $BF_{-0} = 51.33$] which was not seen in DC and SC individuals [$p \geq 0.890$; $BF_{-0} \leq 0.34$]. Accordingly, the size of the visual VAEc significantly differed between groups [$F(2, 16.09) = 3.91$, $p = 0.041$; $BF_{10} = 0.39$]. Games-Howell post-hoc comparisons confirmed that the visual VAEc in the CC group was significantly larger than in the SC group [$p = 0.030$], although these two groups did not significantly differ from the DC group [both $p \geq 0.691$]. Thus, although all three groups recalibrated auditory localization in response to a constant AV spatial discrepancy, only CC individuals additionally recalibrated their visual localization.

The size of the auditory VAEc was neither significantly correlated with the size of the auditory VE [across groups: $r_s = 0.07$, $p = 0.692$; within groups: $r_s \leq 0.43$, $p \geq 0.654$] nor with the size of the auditory VAEi [across groups: $r_s = 0.27$, $p = 0.139$; within groups: $r_s \leq 0.47$, $p \geq 0.534$]. Similarly, the size of the visual VAEc was not significantly correlated with the size of the visual VE [across groups: $r_s = -0.10$, $p = 0.613$; within groups: $r_s \leq -0.06$, $p \geq 0.450$] (see Figure 4).

The duration of blindness may influence the degree to which (but not whether) audiovisual spatial functions recover

Because logMAR (logarithm of the Minimum Angle of Resolution) visual acuity varied between 0.21 and 1.29 in CC individuals and between -0.05 and 0.63 in DC individuals and was considerably worse in both groups compared to SC individuals who all had normal or corrected-to-normal vision, we tested whether visual acuity was correlated with the size of the VE, VAEi, and VAEc effects (see Figure S7). None of the effects were significantly correlated with visual acuity, neither in the CC group [$p > 0.999$], nor in the DC group [$p \geq 0.396$].

We also verified that all participants were able to reliably localize the A and V stimuli. To this end, we computed localization precision in Block 1 as the individual SDs of the localization responses in A and V trials, respectively (see Figure 6). There were no significant group differences in unimodal localization precision between CC, DC, and SC individuals, neither for A precision [$F(2, 17.81) = 1.14$, $p = 0.344$; $BF_{10} = 0.48$] nor for V precision [$F(2, 15.07) = 2.90$, $p = 0.086$; $BF_{10} = 3.00$]. In all three groups, V precision was significantly higher than A precision [$p \leq 0.025$; $BF_{+0} \geq 3.35$]. This suggests that all participants, even those with severe visual impairments, were well able to localize the A and V stimuli used in the present study.

Finally, in an exploratory analysis we compared the sizes of VE, VAEi, and VAEc effects between those CC individuals ($n = 4$) who had their cataract removed relatively early in life (between 5 and 25 months of age) and those CC individuals ($n = 6$) who had cataract surgery only much later in life (at or after 14 years of age). The auditory VE, VAEi, and VAEc tended to be weaker and the visual VE and VAEc tended to be stronger in

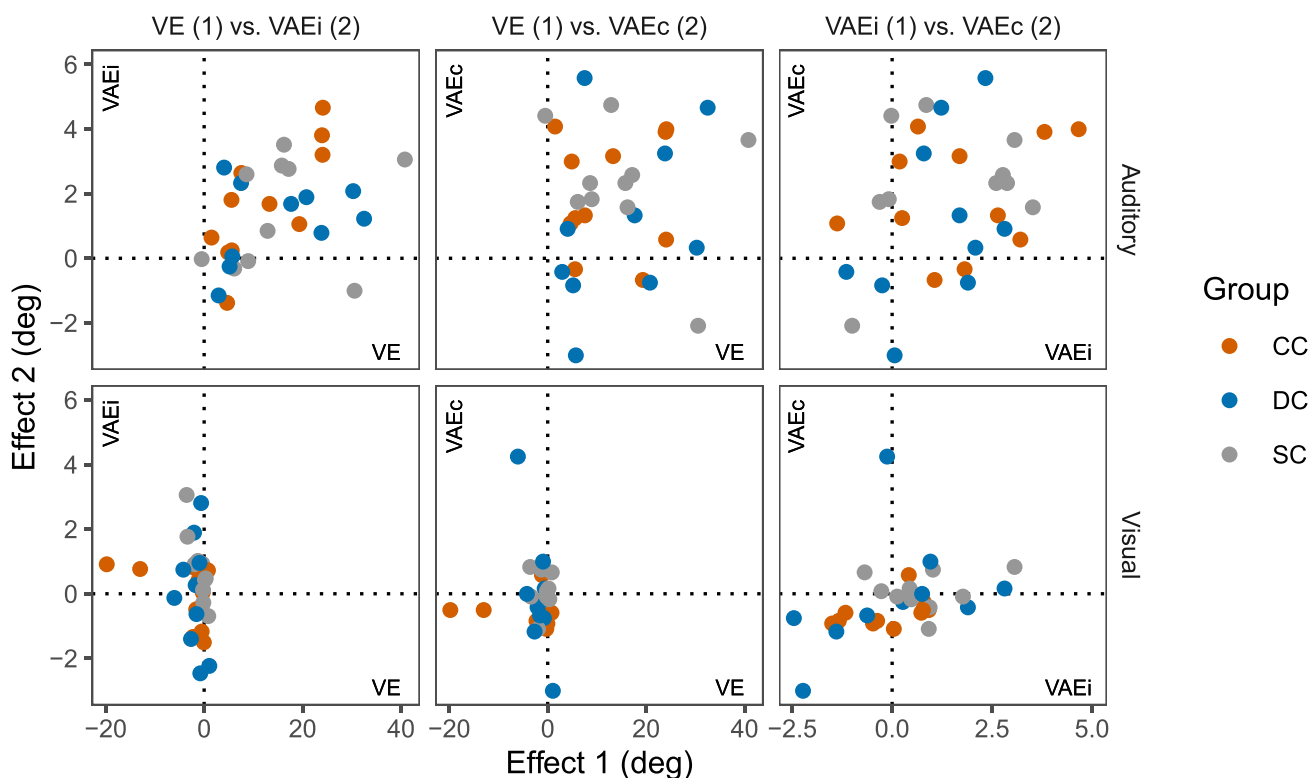


Figure 4. Correlations between ventriloquism effects (VE), immediate (VAEi) and cumulative (VAEc) ventriloquism aftereffects
Bracketed numbers indicate which effect is shown on the x-axis (Effect 1) and which effect is shown on the y-axis (Effect 2) in each subplot. Scatterplots are shown separately for auditory and visual effects. Colors indicate participant groups.

CC individuals who had experienced an extended period of blindness from birth (see Figure S8). However, none of these differences reached statistical significance [$p \geq 0.334$; $BF_{10} \leq 0.66$].

DISCUSSION

Although studies in owls have demonstrated a crucial role of crossmodal experience for adult multisensory learning (recalibration) capabilities, it has remained unclear how early visual experience affects human multisensory learning. In a rare human model, sight-recovery individuals with a (in several cases extensive) history of congenital blindness because of bilateral dense cataracts, the present study found surprisingly normal multisensory spatial integration (as indicated by the VE) and recalibration (as indicated by the auditory VAEi and VAEc) capabilities. However, in addition to using visual spatial information to recalibrate auditory localization (auditory VAEc), as seen in normally sighted controls and individuals who had received cataract removal surgery for incomplete or later developed cataracts, only CC reversal individuals used auditory spatial information to recalibrate visual localization (visual VAEc). Both effects, additional use of auditory input to calibrate visual spatial representations (visual VAEc) and intact multisensory integration as assessed by the VE and computational modeling, were consistently seen across CC individuals who all had received cataract-removal surgery late (i.e., at or after the age of 5 months) but whose individual ages at cataract removal surgery varied considerably up to 33 years. Thus, we speculate that the additional recalibration mechanism observed in CC reversal individuals originates from a neural mechanism existing as a consequence of congenital visual deprivation.

Studies in owls (Brainard and Knudsen, 1998; DeBello et al., 2001; Hyde and Knudsen, 2002) that had manipulated the crossmodal spatial correspondence of auditory and visual input through the use of prisms, found both a reduced elimination of anatomical connections during development and an additional axonal growth that served the atypical mapping induced by prism experience (DeBello et al., 2001). These anatomical changes allowed the prism-reared animals as adults to toggle between two parallel and independent neural organizations, a typical and an atypical one, likely by employing inhibitory mechanisms

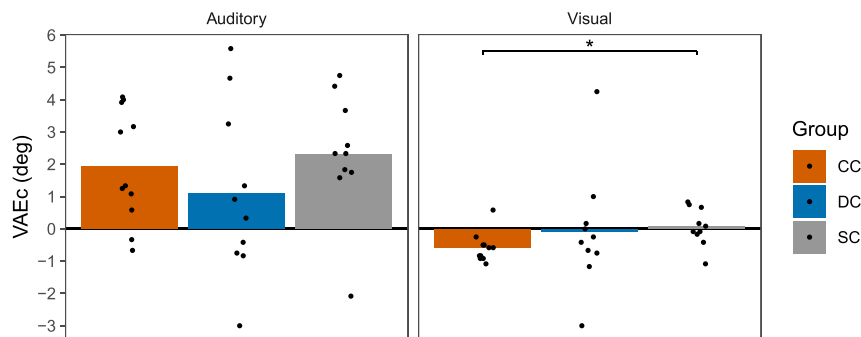


Figure 5. Cumulative ventriloquism aftereffects (VAEc) in CC, DC, and SC individuals

The differences in localization responses (in degrees) between unimodal trials (A and V trials, respectively) in Block 2, in which AV trials featured a constant spatial discrepancy of 10°, and unimodal localization responses in Block 1, in which AV trials had a mean spatial discrepancy of 0°, are shown separately for each group and modality. Bars indicate group averages and dots indicate individual values. Positive values indicate localization biases toward the V location and negative values indicate localization biases toward the A location. A Welch ANOVA indicated a significant group difference in the size of the visual VAEc, which was followed up with Games-Howell post-hoc comparisons. Stars denote significant post-hoc comparisons; * $p < 0.05$. (See also [Figures S2–S4, S7, and S8](#)).

(Hyde and Knudsen, 2002). These findings suggest a more general developmental principle: Experience shapes structural brain networks only during sensitive periods and these neural networks then build the scaffold for future learning. Applied to our human model, in which we compared the effects of presence versus absence of visual input during early development, we speculate that CC individuals had preserved typical multisensory connectivity, allowing them to regain typical multisensory spatial functions, including the VE and the auditory VAEc, after sight restoration. Thus, we suggest that CC individuals were able to “fall back” to a typical organization (Brainard and Knudsen, 1998). In addition, we suggest that they had stabilized an additional set of neural connectivity as a consequence of congenital visual deprivation resulting in the later use of auditory spatial information for calibrating visual representations (visual VAEc). Thus, CC individuals might achieve internal consistency between auditory and visual spatial representations by recalibrating both auditory and visual rather than only auditory spatial representations. A similar coexistence of typical and atypical multisensory links has been observed for the interaction of auditory and visual motion processing (Guerreiro et al., 2016) as well. Although a change of auditory motion perception (a motion aftereffect) after observing visual motion stimuli was found in both sighted controls and CC individuals, a visual after-effect following the presentation of auditory motion was only observed in CC individuals. Thus, in both cases the typical multisensory function characterized by a visual dominance recovered in CC individuals, despite the maintenance of an additional atypical multisensory link.

In congenitally blind humans, evidence for a stronger auditory-driven activity of multisensory parietal cortex than in sighted individuals was reported during auditory localization tasks (Gougoux et al., 2005; Renier et al., 2010; Röder et al., 1999). A stronger auditory influence on multisensory brain regions might either arise from a lack of pruning or a lack of inhibition of exuberant crossmodal connections, as is typically occurring during development (Johannsen and Röder, 2014; see Lewkowicz and Röder, 2012), or from an elaboration of additional connections strengthening the auditory influence in multisensory structures (Rauschecker, 1995). Given the lack of visual information, such a strengthening might be useful for an efficient orienting to and localization of external events. After cataract removal, the additional connections serving auditory spatial input during blindness might (as the atypical connectivity in owls, see Brainard and Knudsen, 1998; DeBello et al., 2001) not be lost but rather be kept and used as a second mechanism for achieving crossmodal consistency. Results from a recent analysis of blood-oxygen-level-dependent (BOLD) resting state activity in CC reversal individuals, most of them with long visual deprivation periods, have indeed suggested altered audio-visual system interactions (Raczy et al., in press). Moreover, previous studies in CC reversal individuals had suggested an enhanced auditory influence in visual cortex after sight restoration (Collignon et al., 2015; Guerreiro et al., 2015). However, how these changes in crossmodal interactions relate to the current behavioral findings remains undissolved.

In contrast to cumulative crossmodal recalibration (VAEc), we did not observe any significant differences in multisensory integration (VE) capabilities between CC individuals and normally sighted controls: CC

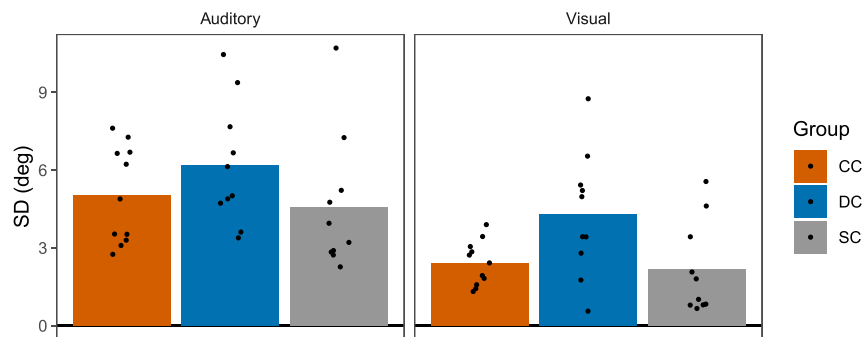


Figure 6. Unimodal localization precision in CC, DC, and SC individuals

The standard deviations of the localization responses (in degrees) in A and V trials of Block 1 are shown separately for each group and modality. Bars indicate group averages and dots indicate individual values.

individuals featured a significant auditory and a nonsignificant visual VE as did the SC group. A strong visual influence on auditory localization in CC individuals fits with their higher visual compared to their auditory localization precision in the present study, as found in the normally sighted controls. Computational modeling analyses indeed suggested an employment of similar multisensory integration principles in CC individuals as typically observed in the normally sighted population (Körding et al., 2007; Wozny et al., 2010) and replicated in the present study in both SC and DC individuals: Causal inference models explained multisensory integration best in all three groups. Consistent with previous studies, there was some variation in the best-fitting CI decision strategy across individuals. Nevertheless, the majority of our participants, including the majority of the CC individuals, were best described by a probability matching strategy (Wozny et al., 2010). Regardless of the variation in individual decision strategies, the superiority of the CI models indicates that the behavioral localization responses were based on a normative causal inference process in all three groups. Thus, in a task that is typically characterized by a strong visual dominance and by a decline of the visual influence over development (Rohlf et al., 2020), we observed extensive recovery in individuals who had experienced a phase of congenital blindness before sight restoration (see also Senna et al., 2021). A similar result pattern emerged for immediate (trial-by-trial) recalibration: Despite an additional visual VAEc in CC individuals, an additional visual VAEi was not detected, whereas a typical auditory VAEi was revealed. In fact, the size of the auditory VAEi and VE was significantly correlated in the CC group as has been previously demonstrated in normally sighted individuals (Rohlf et al., 2021). Thus, the present results provide further evidence for a dissociation of cumulative crossmodal spatial recalibration (VAEc) and both immediate crossmodal recalibration (VAEi) and multisensory integration (VE), suggesting that the latter two processes originate from overlapping neural mechanisms (Park and Kayser, 2019).

The results for the VE contrast with previous reports of an impaired integration of visual and auditory speech signals in CC individuals (Putzar et al., 2007, 2010). However, we think there is a crucial difference between audio-visual speech perception and audio-visual localization: In contrast to localization, speech perception is better achieved through the auditory than through the visual (lipreading) channel. Prospective studies have shown that the McGurk effect, that is, the fusion of incongruent auditory and visual speech signals, increases between 3 and 9 years of age (Hirst et al., 2018). Younger children more likely reported the auditory input and needed a relatively more reliable visual stimulus in order to report a fused percept (Hirst et al., 2018). By contrast, the VE was stronger in children than in adults (Rohlf et al., 2020), that is, the VE declines rather than increases with age. Thus, in both cases more adult-like multisensory percepts arise across development from an increased weighting of the less reliable (less dominant) sensory input. This requires an increased weighting of vision in audio-visual speech perception, but an increased weighting of the auditory input for audio-visual localization (and hence a reduction of the visual influence). In CC individuals, the latter is likely already in place but the first can only be developed after sight restoration. Thus, the different developmental time course for audiovisual speech and spatial processes might have favored recovery of multisensory spatial integration despite the remaining and often quite severe post-surgical visual impairments typically seen in CC individuals. Our modeling results suggest that CC individuals weighted auditory and visual spatial information according to their current relative reliability similar to normally sighted individuals. A typical VE in CC individuals does, however, not necessarily mean that the VE

does not require visual or crossmodal input to emerge, but rather that this visual or crossmodal input is not required during the first phase of life (Senna et al., 2021).

These results are seemingly at odds with studies in dark reared cats in which no multisensory enhancement was found in multisensory neurons of the superior colliculus (Wallace et al., 2004; Yu et al., 2019). Recently, however, multisensory gains in superior colliculus neurons have been observed when the animals were tested in light rather than in the dark (Smyre et al., 2021). We conducted our experiment in a lit room and, moreover, unlike the previously tested dark-reared cats, the CC individuals of the present study had extensive experience with audio-visual stimuli in natural environments after cataract-removal surgery, many for several years. This experience might have allowed for a recovery of the typical visual influence on auditory localization. A prerequisite for such a recovery is that unisensory visual spatial representations recover after sight restoration. In fact, a regular visual topographic organization has been demonstrated in visually deprived owls (Du Lac and Knudsen, 1991) and ferrets (King and Carlile, 1993) that were raised with binocular lid suture, and for primary visual cortex in human CC individuals (Sourav et al., 2018). Moreover, both in owls (Knudsen et al., 1991) and ferrets (King and Carlile, 1993) at least a crude alignment of auditory and visual receptive fields was observed immediately after terminating visual deprivation. It is possible that diffuse residual light perception, which exists even in the presence of bilateral dense cataracts similarly as in lid-sutured animal models (Sherman and Spear, 1982), had contributed to the preservation of the typical crossmodal connectivity in the CC individuals of the present study.

Why did recovery in CC individuals differ between multisensory integration (VE) and recalibration (VAEc)? There is multiple independent evidence suggesting that multisensory integration (VE) and crossmodal recalibration (VAEc) reflect at least partially distinct mechanisms: Event-related potential studies revealed an early versus late modulation for the VAEc and VE, respectively (Bonath et al., 2007; Bruns et al., 2011; Bruns and Röder, 2010). Prospective studies in children have recently observed that the VE emerges prior to the VAEc (Rohlf et al., 2020, 2021), thus providing developmental evidence for a dissociation of both effects. Moreover, the VAEc was found to be independent of cue reliability whereas the VE, in accord with an extensive literature (Alais and Burr, 2019), varied with cue reliability even in the youngest (5 years old) children tested (Rohlf et al., 2020, 2021). Zaidel et al. (2011) suggested that crossmodal recalibration uses a fixed-ratio weighting of the crossmodal cues which is independent of current cue reliability (but see recent results from Hong et al., 2021). Although multisensory integration aims at improving precision and, thus, should take cue reliability into account, crossmodal recalibration might aim at achieving internal consistency (i.e., accuracy) of individual cues which does not depend on reliability (Zaidel et al., 2011).

If crossmodal recalibration indeed uses a fixed-ratio weighting of cues, the question arises how this ratio is set. The result that crossmodal recalibration emerges relatively late (>7 years) in development (Rohlf et al., 2020) suggests an experience-dependence. Recent studies have provided evidence that crossmodal temporal biases with a high intra-individual stability are altered in CC individuals, suggesting that crossmodal settings are indeed acquired and stabilized during a sensitive phase of development (Chen et al., 2017; Badde et al., 2020). Similarly, the fixed-ratio weighting of sensory cues during crossmodal recalibration seems to have a high stability and consistency across humans and other primates (Zaidel et al., 2011) and, thus, might be stabilized based on experience during a sensitive period. This is supported by the high inter-individual consistency with which the visual VAEc emerged in the CC group. However, it should be considered that in the present study a significant group difference in the visual VAEc was only obtained between the CC and SC but not between the CC and DC groups.

In summary, we found multisensory integration to be indistinguishable across groups (no visual VE emerged in the CC group) whereas for cumulative crossmodal calibration we observed an expansion to recalibrating visual localization by auditory input in the CC group. We speculate that the auditory VE, which seems to emerge early in development (Rohlf et al., 2020) and is highly dependent on cue reliability both in children (Rohlf et al., 2020, 2021) and in adults (Alais and Burr, 2004), indicates a multisensory process which has to maintain a high level of plasticity throughout life to be able to instantly adjust to changes in the environment because of the need to maximize localization precision. By contrast, crossmodal recalibration is mainly important for maintaining localization accuracy. After the body has reached a certain level of maturity, expected future adaptations are relatively minor for the majority of the population (similar as for crossmodal temporal biases, Badde et al., 2020). If experience is expected to remain stable in the population after a certain developmental phase, a sensitive period is an efficient evolutionary strategy (Frankenhuis and Walasek, 2020).

Limitations of the study

Our results demonstrate that early crossmodal experience is not necessary for the emergence of normal multisensory spatial integration (VE) and crossmodal recalibration (auditory VAEc) capabilities, but that lack of vision during early development results in the establishment of additional recalibration mechanisms (visual VAEc) not seen in sighted individuals or individuals with a later developed visual impairment. Because of the rare patient group tested in our study, our statistical analyses had to be based on a relatively small sample size. However, the typical pattern of multisensory integration and learning (auditory VE, VAEi, and VAEc) was replicated in all three groups of the present study, and the additional visual VAEc observed in the CC group was highly consistent across CC individuals. All these effects robustly emerged despite the use of a dual-task paradigm in which participants had to report both the auditory and the visual location in crossmodal trials and which might, thus, have diminished integration of the crossmodal cues compared to a joint report. Our behavioral findings can, however, not provide direct insights into the possible neural basis for the change in recalibration mechanisms resulting from a lack of early visual experience. Besides, we are not able to precisely delineate the timing of the sensitive period because we had to rely on naturally occurring cases of blindness and sight recovery in our human model. We applied stringent selection criteria for CC individuals based on medical records to make total and dense cataract at birth likely. Nevertheless, some uncertainty about the precise morphology of the cataractous lens inevitably remains in patients who presented for surgery long after birth. In addition, although the age of cataract onset is known in CC individuals (congenital), it is not clearly defined in DC individuals because of the typically gradual onset of developmental cataracts. However, because age at surgery varied considerably but results were quite consistent across CC individuals (see [Figure S8](#)), our findings point to a crucial role of the first year in human brain development in accord with neuroanatomical studies ([Gilmore et al., 2020](#)).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - General procedure
 - Block 1 (VE and VAEi)
 - Block 2 (VAEc)
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - VE, VAEi, and VAEc
 - Model descriptions
 - Model Parameters and Simulations
 - Model fitting and comparison

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104439>.

ACKNOWLEDGMENTS

This research was funded by grants from the German Research Foundation (DFG TRR 169/A1 to P.B. and B.R. and DFG Ro 2625/10-1 to B.R.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, P.B. and B.R.; Methodology, P.B.; Formal Analysis, P.B. and L.L.; Investigation, M.J.S.G., I.S., and S.S.R.; Resources, R.K.; Data Curation, K.P.; Writing – Original Draft, P.B., L.L., and B.R.; Writing – Review & Editing, all authors; Visualization, P.B. and L.L.; Supervision, R.K. and B.R.; Funding Acquisition, P.B. and B.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 5, 2021

Revised: February 14, 2022

Accepted: May 6, 2022

Published: June 17, 2022

REFERENCES

- Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>.
- Alais, D., and Burr, D. (2019). Cue combination within a bayesian framework. In *Multisensory Processes: The Auditory Perspective*, A.K.C. Lee, M.T. Wallace, A.B. Coffin, A.N. Popper, and R.R. Fay, eds. (Springer), pp. 9–31. https://doi.org/10.1007/978-3-030-10461-0_2.
- Azanon, E., Camacho, K., Morales, M., and Longo, M.R. (2018). The sensitive period for tactile remapping does not include early infancy. *Child. Dev.* 89, 1394–1404. <https://doi.org/10.1111/cdev.12813>.
- Badde, S., Ley, P., Rajendran, S.S., Shareef, I., Kekunnaya, R., and Röder, B. (2020). Sensory experience during early sensitive periods shapes cross-modal temporal biases. *Elife* 9, e61238. <https://doi.org/10.7554/eLife.61238>.
- Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., and Hillyard, S.A. (2007). Neural basis of the ventriloquist illusion. *Curr. Biol.* 17, 1697–1703. <https://doi.org/10.1016/j.cub.2007.08.050>.
- Brainard, M.S., and Knudsen, E.I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci.* 18, 3929–3942. <https://doi.org/10.1523/JNEUROSCI.18-10-03929.1998>.
- Bruns, P., Liebnau, R., and Röder, B. (2011). Cross-modal training induces changes in spatial representations early in the auditory processing pathway. *Psychol. Sci.* 22, 1120–1126. <https://doi.org/10.1177/0956797611416254>.
- Bruns, P., and Röder, B. (2010). Tactile capture of auditory localization: an event-related potential study. *Eur. J. Neurosci.* 31, 1844–1857. <https://doi.org/10.1111/j.1460-9568.2010.07232.x>.
- Bruns, P., and Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Sci. Rep.* 5, 12739. <https://doi.org/10.1038/srep12739>.
- Byrd, R.H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* 16, 1190–1208. <https://doi.org/10.1137/0916069>.
- Cao, Y., Summerfield, C., Park, H., Giordano, B.L., and Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron* 102, 1076–1087.e8. <https://doi.org/10.1016/j.neuron.2019.03.043>.
- Chen, Y.-C., Lewis, T.L., Shore, D.I., and Maurer, D. (2017). Early binocular input is critical for development of audiovisual but not visuotactile simultaneity perception. *Curr. Biol.* 27, 583–589. <https://doi.org/10.1016/j.cub.2017.01.009>.
- Collignon, O., Charbonneau, G., Lassonde, M., and Lepore, F. (2009). Early visual deprivation alters multisensory processing in peripersonal space. *Neuropsychologia* 47, 3236–3243. <https://doi.org/10.1016/j.neuropsychologia.2009.07.025>.
- Collignon, O., Dormal, G., de Heering, A., Lepore, F., Lewis, T.L., and Maurer, D. (2015). Long-lasting crossmodal cortical reorganization triggered by brief postnatal visual deprivation. *Curr. Biol.* 25, 2379–2383. <https://doi.org/10.1016/j.cub.2015.07.036>.
- DeBello, W.M., Feldman, D.E., and Knudsen, E.I. (2001). Adaptive axonal remodeling in the midbrain auditory space map. *J. Neurosci.* 21, 3161–3174. <https://doi.org/10.1523/JNEUROSCI.21-09-03161.2001>.
- De Heering, A., Dormal, G., Pelland, M., Lewis, T., Maurer, D., and Collignon, O. (2016). A brief period of postnatal visual deprivation alters the balance between auditory and visual attention. *Curr. Biol.* 26, 3101–3105. <https://doi.org/10.1016/j.cub.2016.10.014>.
- De Winkel, K.N., Katliar, M., and Bühlhoff, H.H. (2015). Forced fusion in multisensory heading estimation. *PLoS. One.* 10, e0127104. <https://doi.org/10.1371/journal.pone.0127104>.
- Du Lac, S., and Knudsen, E.I. (1991). Early visual deprivation results in a degraded motor map in the optic tectum of barn owls. *Proc. Natl. Acad. Sci. U.S.A.* 88, 3426–3430. <https://doi.org/10.1073/pnas.88.8.3426>.
- Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. <https://doi.org/10.1038/415429a>.
- Feldman, D.E., and Knudsen, E.I. (1997). An anatomical basis for visual calibration of the auditory space map in the barn owl's midbrain. *J. Neurosci.* 17, 6820–6837. <https://doi.org/10.1523/JNEUROSCI.17-17-06820.1997>.
- Frankenhuis, W.E., and Walasek, N. (2020). Modeling the evolution of sensitive periods. *Dev. Cogn. Neurosci.* 41, 100715. <https://doi.org/10.1016/j.dcn.2019.100715>.
- Frissen, I., Vroomen, J., de Gelder, B., and Bertelson, P. (2003). The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychol.* 113, 315–327. [https://doi.org/10.1016/S0001-6918\(03\)00043-X](https://doi.org/10.1016/S0001-6918(03)00043-X).
- Fryer, L., Freeman, J., and Pring, L. (2014). Touching words is not enough: how visual experience influences haptic-auditory associations in the “Bouba-Kiki” effect. *Cognition* 132, 164–173. <https://doi.org/10.1016/j.cognition.2014.03.015>.
- Geisler, W.S. (2011). Contributions of ideal observer theory to vision research. *Vision. Res.* 51, 771–781. <https://doi.org/10.1016/j.visres.2010.09.027>.
- Gilmore, J.H., Langworthy, B., Girault, J.B., Fine, J., Jha, S.C., Kim, S.H., Cornea, E., and Styner, M. (2020). Individual variation of human cortical structure is established in the first year of life. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging.* 5, 971–980. <https://doi.org/10.1016/j.bpsc.2020.05.012>.
- Gougoux, F., Zatorre, R.J., Lassonde, M., Voss, P., and Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS. Biol.* 3, e27. <https://doi.org/10.1371/journal.pbio.0030027>.
- Guerreiro, M.J.S., Putzar, L., and Röder, B. (2015). The effect of early visual deprivation on the neural bases of multisensory processing. *Brain* 138, 1499–1504. <https://doi.org/10.1093/brain/awv076>.
- Guerreiro, M.J.S., Putzar, L., and Röder, B. (2016). Persisting cross-modal changes in sight-recovery individuals modulate visual perception. *Curr. Biol.* 26, 3096–3100. <https://doi.org/10.1016/j.cub.2016.08.069>.
- Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefanczyk, M., Ward, J., and Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition* 175, 114–121. <https://doi.org/10.1016/j.cognition.2018.02.023>.
- Hirst, R.J., Stacey, J.E., Cragg, L., Stacey, P.C., and Allen, H.A. (2018). The threshold for the McGurk effect in audio-visual noise decreases with development. *Sci. Rep.* 8, 12372. <https://doi.org/10.1038/s41598-018-30798-8>.
- Hong, F., Badde, S., and Landy, M.S. (2021). Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. *PLoS. Comput. Biol.* 17, e1008877. <https://doi.org/10.1371/journal.pcbi.1008877>.
- Hyde, P.S., and Knudsen, E.I. (2002). The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map. *Nature* 415, 73–76. <https://doi.org/10.1038/415073a>.

- Jackson, C.V. (1953). Visual factors in auditory localization. *Q. J. Exp. Psychol.* 5, 52–65. <https://doi.org/10.1080/17470215308416626>.
- Johannsen, J., and Röder, B. (2014). Uni- and crossmodal refractory period effects of event-related potentials provide insights into the development of multisensory processing. *Front. Hum. Neurosci.* 8, 552. <https://doi.org/10.3389/fnhum.2014.00552>.
- Kayser, C., and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol.* 13, e1002075. <https://doi.org/10.1371/journal.pbio.1002075>.
- King, A.J., and Carlile, S. (1993). Changes induced in the representation of auditory space in the superior colliculus by rearing ferrets with binocular eyelid suture. *Exp. Brain Res.* 94, 444–455. <https://doi.org/10.1007/BF00230202>.
- Knudsen, E.I. (1998). Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science* 279, 1531–1533. <https://doi.org/10.1126/science.279.5356.1531>.
- Knudsen, E.I., Esterly, S.D., and du Lac, S. (1991). Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls; acoustic basis and behavioral correlates. *J. Neurosci.* 11, 1727–1747. <https://doi.org/10.1523/JNEUROSCI.11-06-01727.1991>.
- Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS One.* 2, e943. <https://doi.org/10.1371/journal.pone.0000943>.
- Lewkowicz, D.J., and Röder, B. (2012). Development of multisensory processing and the role of early experience. In *The New Handbook of Multisensory Processes*, B.E. Stein, ed. (MIT Press), pp. 607–626.
- Ley, P., Bottari, D., Shenoy, B.H., Kekunnaya, R., and Röder, B. (2013). Partial recovery of visual-spatial remapping of touch after restoring vision in a congenitally blind man. *Neuropsychologia* 51, 1119–1123. <https://doi.org/10.1016/j.neuropsychologia.2013.03.004>.
- Li, L., Rehr, R., Bruns, P., Gerkmann, T., and Röder, B. (2020). A survey on probabilistic models in human perception and machines. *Front. Robot. AI.* 7, 85. <https://doi.org/10.3389/frobt.2020.00085>.
- Linkenhoker, B.A., von der Ohe, C.G., and Knudsen, E.I. (2005). Anatomical traces of juvenile learning in the auditory system of adult barn owls. *Nat. Neurosci.* 8, 93–98. <https://doi.org/10.1038/nn1367>.
- Mohl, J.T., Pearson, J.M., and Groh, J.M. (2020). Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. *J. Neurophysiol.* 124, 715–727. <https://doi.org/10.1152/jn.00046.2020>.
- Neath, A.A., and Cavanaugh, J.E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdiscip. Rev. Comput. Stat.* 4, 199–203. <https://doi.org/10.1002/wics.199>.
- Pant, R., Guerreiro, M.J.S., Ley, P., Bottari, D., Shareef, I., Kekunnaya, R., and Röder, B. (2021). The size-weight illusion is unimpaired in individuals with a history of congenital visual deprivation. *Sci. Rep.* 11, 6693. <https://doi.org/10.1038/s41598-021-86227-w>.
- Park, H., and Kayser, C. (2019). Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. *Elife* 8, e47001. <https://doi.org/10.7554/eLife.47001>.
- Park, H., and Kayser, C. (2021). The neurophysiological basis of the trial-wise and cumulative ventriloquism aftereffects. *J. Neurosci.* 41, 1068–1079. <https://doi.org/10.1523/JNEUROSCI.2091-20.2020>.
- Park, H., Nannt, J., and Kayser, C. (2021). Sensory- and memory-related drivers for altered ventriloquism effects and aftereffects in older adults. *Cortex* 135, 298–310. <https://doi.org/10.1016/j.cortex.2020.12.001>.
- Putzar, L., Goerendt, I., Lange, K., Rösler, F., and Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nat. Neurosci.* 10, 1243–1245. <https://doi.org/10.1038/nn1978>.
- Putzar, L., Gondan, M., and Röder, B. (2012). Basic multisensory functions can be acquired after congenital visual pattern deprivation in humans. *Dev. Neuropsychol.* 37, 697–711. <https://doi.org/10.1080/87565641.2012.696756>.
- Putzar, L., Hötting, K., and Röder, B. (2010). Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restor. Neurol. Neurosci.* 28, 251–257. <https://doi.org/10.3233/RNN-2010-0526>.
- Raczky, K., Hölig, C., Guerreiro, M.J.S., Lingareddy, S., Kekunnaya, R., and Röder, B. Typical resting state activity of the brain requires visual input during an early sensitive period. *Brain. Commun.* <https://doi.org/10.1101/2021.06.09.446724>.
- Raftery, A.E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25, 111–163. <https://doi.org/10.2307/271063>.
- Rauschecker, J.P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* 18, 36–43. [https://doi.org/10.1016/0166-2236\(95\)93948-W](https://doi.org/10.1016/0166-2236(95)93948-W).
- Renier, L.A., Anurova, I., De Volder, A.G., Carlson, S., VanMeter, J., and Rauschecker, J.P. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron* 68, 138–148. <https://doi.org/10.1016/j.neuron.2010.09.021>.
- Röder, B., Kusmirek, A., Spence, C., and Schicke, T. (2007). Developmental vision determines the reference frame for the multisensory control of action. *Proc. Natl. Acad. Sci. U S A.* 104, 4753–4758. <https://doi.org/10.1073/pnas.0607158104>.
- Röder, B., Rösler, F., and Spence, C. (2004). Early vision impairs tactile perception in the blind. *Curr. Biol.* 14, 121–124. <https://doi.org/10.1016/j.cub.2003.12.054>.
- Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S.A., and Neville, H.J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. <https://doi.org/10.1038/22106>.
- Rohde, M., van Dam, L.C.J., and Ernst, M.O. (2016). Statistically optimal multisensory cue integration: a practical tutorial. *Multisens. Res.* 29, 279–317. <https://doi.org/10.1163/22134808-00002510>.
- Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* 13, e1002073. <https://doi.org/10.1371/journal.pbio.1002073>.
- Rohlf, S., Bruns, P., and Röder, B. (2021). The effects of cue reliability on crossmodal recalibration in adults and children. *Multisens. Res.* 34, 743–761. <https://doi.org/10.1163/22134808-bja10053>.
- Rohlf, S., Li, L., Bruns, P., and Röder, B. (2020). Multisensory integration develops prior to crossmodal recalibration. *Curr. Biol.* 30, 1726–1732.e7. <https://doi.org/10.1016/j.cub.2020.02.048>.
- Schorr, E.A., Fox, N.A., van Wassenhove, V., and Knudsen, E.I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proc. Natl. Acad. Sci. U S A.* 102, 18748–18750. <https://doi.org/10.1073/pnas.0508862102>.
- Senna, I., Andres, E., McKyton, A., Ben-Zion, I., Zohary, E., and Ernst, M.O. (2021). Development of multisensory integration following prolonged early-onset visual deprivation. *Curr. Biol.* 31, 4879–4885.e6. <https://doi.org/10.1016/j.cub.2021.08.060>.
- Sherman, S.M., and Spear, P.D. (1982). Organization of visual pathways in normal and visually deprived cats. *Physiol. Rev.* 62, 738–855. <https://doi.org/10.1152/physrev.1982.62.2.738>.
- Smyre, S.A., Wang, Z., Stein, B.E., and Rowland, B.A. (2021). Multisensory enhancement of overt behavior requires multisensory experience. *Eur. J. Neurosci.* 54, 4514–4527. <https://doi.org/10.1111/ejn.15315>.
- Sourav, S., Bottari, D., Kekunnaya, R., and Röder, B. (2018). Evidence of a retinotopic organization of early visual cortex but impaired extrastriate processing in sight recovery individuals. *J. Vis.* 18, 22. <https://doi.org/10.1167/18.3.22>.
- Sourav, S., Kekunnaya, R., Shareef, I., Banerjee, S., Bottari, D., and Röder, B. (2019). A protracted sensitive period regulates the development of cross-modal sound-shape associations in humans. *Psychol. Sci.* 30, 1473–1482. <https://doi.org/10.1177/0956797619866625>.
- Stawicki, M., Majdak, P., and Baskent, D. (2019). Ventriloquist illusion produced with virtual acoustic spatial cues and asynchronous audiovisual stimuli in both young and older individuals. *Multisens. Res.* 32, 745–770. <https://doi.org/10.1163/22134808-20191430>.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, O.F., Dropmann, D., Boutin, B., et al. (2018). Bayesian inference for psychology. Part II: example applications with JASP. *Psychon. Bull.*

Rev. 25, 58–76. <https://doi.org/10.3758/s13423-017-1323-7>.

Wallace, M.T., Perrault, T.J., Jr., Hairston, W.D., and Stein, B.E. (2004). Visual experience is necessary for the development of multisensory integration. *J. Neurosci.* 24, 9580–9584. <https://doi.org/10.1523/JNEUROSCI.2535-04.2004>.

Wallace, M.T., and Stein, B.E. (2007). Early experience determines how the senses will interact. *J. Neurophysiol.* 97, 921–926. <https://doi.org/10.1152/jn.00497.2006>.

Watson, D.M., Akeroyd, M.A., Roach, N.W., and Webb, B.S. (2019). Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history. *Sci. Rep.* 9, 8513. <https://doi.org/10.1038/s41598-019-44984-9>.

World Health Organization. (2019). International Classification of Diseases. (11th revision). <https://icd.who.int/>.

Wozny, D.R., Beierholm, U.R., and Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Comput. Biol.* 6, e1000871. <https://doi.org/10.1371/journal.pcbi.1000871>.

Wozny, D.R., and Shams, L. (2011). Computational characterization of visually induced auditory spatial adaptation. *Front. Integr. Neurosci.* 5, 75. <https://doi.org/10.3389/fnint.2011.00075>.

Yu, L., Cuppini, C., Xu, J., Rowland, B.A., and Stein, B.E. (2019). Cross-modal competition: the default computation for multisensory processing. *J. Neurosci.* 39, 1374–1385.

<https://doi.org/10.1523/JNEUROSCI.1806-18.2018>.

Zaidel, A., Turner, A.H., and Angelaki, D.E. (2011). Multisensory calibration is independent of cue reliability. *J. Neurosci.* 31, 13949–13962. <https://doi.org/10.1523/JNEUROSCI.2732-11.2011>.

Zierul, B., Röder, B., Tempelmann, C., Bruns, P., and Noesselt, T. (2017). The role of auditory cortex in the spatial ventriloquism aftereffect. *Neuroimage* 162, 257–268. <https://doi.org/10.1016/j.neuroimage.2017.09.002>.

Zierul, B., Tong, J., Bruns, P., and Röder, B. (2019). Reduced multisensory integration of self-initiated stimuli. *Cognition* 182, 349–359. <https://doi.org/10.1016/j.cognition.2018.10.019>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw behavioral data	This paper	https://doi.org/10.25592/uhhfdm.10191
Software and algorithms		
Presentation	Neurobehavioral Systems, Inc.	https://www.neurobs.com/
R version 4.0.4	R Foundation for Statistical Computing	https://www.R-project.org/
JASP version 0.14	JASP Team	https://jasp-stats.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Patrick Bruns (patrick.bruns@uni-hamburg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw behavioral data have been deposited in the research data repository of the University of Hamburg and are publicly available as of the date of publication. The DOI is listed in the key resources table. De-identified subject characteristics are provided in [Table S3](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A group of 11 individuals (aged 12-41 years, $M = 25.9$, $SD = 10.1$, two female and nine male) with a history of dense bilateral congenital cataracts (CC), resulting in a transient period of blindness (according to the classification of the [World Health Organization, 2019](#)) from birth, was recruited. Their cataracts were surgically removed between the ages of 5 months and 33 years ($M = 11.7$ years), with surgery taking place at or before the age of two years in four individuals, at the age of six years in one individual, and at or after the age of 14 years in the remaining six individuals (for individual participant characteristics, see [Table S3](#)). It should be noted that the exact duration of blindness does not necessarily correspond to age at surgery in CC individuals with absorbed cataracts: Two individuals from our CC group who had presented for surgery with absorbed cataracts later in life (>14 years of age) did no longer fulfil the World Health Organization (2019) criteria for blindness (but still had severe vision impairment) at the time of surgery (see [Table S3](#)). However, although the exact age of cataract absorption in these two individuals is unknown to us, cataract absorption typically does not occur before middle childhood, suggesting that they had experienced an extended period of blindness from birth. Cataract history, as well as associated factors such as presence of sensory nystagmus, absence of fundus view prior to surgery, and positive family history, were confirmed by medical records. Data of one additional CC individual who was operated only in one eye at the time of testing were not included in the analyses because of poor vision in the operated eye (logMAR visual acuity of 1.55). The remaining CC individuals had a mean logMAR visual acuity of 0.81 (range: 0.21 to 1.29).

An additional group of 10 individuals (aged 9-28 years, $M = 14.8$, $SD = 5.2$, seven female and three male) who were treated for developmental cataracts (DC) served as a control group (for individual participant characteristics, see [Table S3](#)). The DC individuals had a mean logMAR visual acuity of 0.24 (range: -0.05 to 0.63). A Welch two-sample t test revealed a significant difference in visual acuity between the CC and

the DC group, $t(18.78) = 4.44$, $p < .001$. All CC and DC individuals reported normal hearing and had no known neurological disorders. Cataract-reversal individuals in the CC and DC groups were compared to a group of 10 normally sighted control (SC) individuals (aged 15-30 years, $M = 19.1$, $SD = 4.8$, three female and seven male) who had a typical development of all sensory systems. All 10 SC individuals had normal or corrected-to-normal vision according to their self-report. Additional visual acuity data were available for seven SC individuals who had a mean logMAR acuity of -0.17 (range: -0.24 to -0.10). The age range of the SC group (15-30 years) largely overlapped with the age ranges of the CC and DC groups, except for the youngest participants in the DC group (minimal age of 9 years) and the oldest participants in the CC group (maximal age of 41 years). However, previous studies had shown that the experimental effects of interest, the VE and VAE, were adult-like in 8-9 year-old children (Rohlf et al., 2020) and did not differ between younger and older adults (Stawicki et al., 2019) or only between younger adults and adults aged above 60 years (Park et al., 2021).

All CC, DC, and SC individuals were recruited at the LV Prasad Eye Institute in Hyderabad, India. Participants or, for minors, their legal guardians provided written informed consent prior to taking part in the study. Expenses associated with taking part in the study, such as travel costs, were reimbursed and minors received a small gift. The study was approved by the Local Ethics Committee of the Faculty of Psychology and Human Movement at the University of Hamburg, Germany, as well as by the Institutional Review Board at the LV Prasad Eye Institute in Hyderabad, India, and was performed in accordance with the ethical standards laid down in the Declaration of Helsinki.

METHOD DETAILS

General procedure

The experimental setup resembled those used in typical studies of the VE and VAE (Frissen et al., 2003; Jackson, 1953; Körding et al., 2007; Park and Kayser, 2019, 2021; Rohe and Noppeney, 2015; Wozny et al., 2010; Wozny and Shams, 2011; Zierul et al., 2017). Participants faced the center of four loudspeakers (Companion 2, Bose Corporation, Framingham, MA, USA) which were placed at a distance of 45 cm with eccentricities of $\pm 5^\circ$ and $\pm 15^\circ$ from the participants' straight-ahead position (0°). The loudspeakers were covered by an acoustically transparent cloth. Four pairs of red LEDs were attached on top of the loudspeaker array, one pair at the center of each loudspeaker location. LED pairs rather than single LEDs were used to increase the saliency of the visual stimulation because we were concerned that residual unimodal visual localization impairments in CC individuals (and possibly in DC individuals) would have otherwise masked any genuine group differences in multisensory processing (Badde et al., 2020). For visual stimulation, LEDs were illuminated for 35 ms. Auditory stimuli were white noise bursts with a duration of 35 ms. Audiovisual stimuli were always presented synchronously. Participants indicated perceived stimulus locations with an array of 12 pushbuttons which was placed directly in front of the loudspeakers and which had the same width as the loudspeaker array. A schematic illustration of the experimental setup is shown in Figure 1.

The experiment consisted of two blocks of 240 trials each. Block 1 was designed to measure VE and VAE_i, and Block 2 was designed to measure VAE_c. Each block lasted approximately 15 min and included 40 unimodal auditory (A) trials, 40 unimodal visual (V) trials, and 160 audiovisual (AV) trials. After each trial, participants indicated the perceived location of the A and/or V stimulus by pressing the corresponding button in front of the loudspeaker/LED array. On AV trials, two responses were required, one for the A and one for the V stimulus. Half of the participants reported A first and V second throughout the experiment, and vice versa for the other half of the participants. The task was self-paced and accuracy was stressed over response speed. The next trial was presented between 900 and 1400 ms (randomly determined) after the (second) response.

Block 1 (VE and VAE_i)

A and V trials were equally distributed across the four locations (10 trials per location each). In AV trials, each possible combination of the four A and V locations occurred 10 times, resulting in AV spatial discrepancies of $\pm 30^\circ$ (10 trials each), $\pm 20^\circ$ (20 trials each), $\pm 10^\circ$ (30 trials each), or 0° (40 trials). A, V, and AV trials were presented in a pseudorandom order, with the constraint that A and V trials were equally often preceded (disregarding any interjacent unimodal trials) by AV trials with $\pm 20^\circ$, $\pm 10^\circ$, and 0° spatial discrepancy (8 trials for each preceding AV discrepancy). AV trials with $\pm 30^\circ$ discrepancy were never immediately followed by a unimodal trial.

Block 2 (VAEc)

A and V trials were equally distributed across the four locations as in Block 1. However, in AV trials the V stimulus was always 10° to the right of the A stimulus. Thus, only three combinations of A and V locations were presented (A:-15°, V:-5°, 53 trials; A:-5°, V: 5°, 54 trials; A: 5°, V: 15°, 53 trials). A, V, and AV trials were presented in random order.

QUANTIFICATION AND STATISTICAL ANALYSIS

VE, VAEi, and VAEc

As in previous studies using categorical localization responses (Bruns et al., 2011; Bruns and Röder, 2010; Zierul et al., 2017), individual responses were coded as 1-12 (corresponding to the 12 pushbuttons from left to right) and then averaged across the 10 unimodal trials (A and V separately) per location from Block 1. These baseline localization scores were then subtracted from localization responses in each trial of Block 1 and 2 to determine signed localization errors (negative indicates leftward and positive indicates rightward localization bias). Localization errors were converted from response units to degrees for visualization purposes.

Ventriloquism effects (VE) were calculated as difference in localization errors between AV trials with a rightward discrepancy (V to the right from A location) and AV trials with a leftward discrepancy (V to the left from A location) in Block 1. Immediate ventriloquism aftereffects (VAEi) were calculated as difference in localization errors between unimodal trials which were preceded by an AV trial with a rightward discrepancy and unimodal trials which were preceded by an AV trial with a leftward discrepancy in Block 1. Cumulative ventriloquism aftereffects (VAEc) were calculated as mean localization errors in unimodal trials of Block 2, which indicate the difference in localization responses between Block 1 (average AV discrepancy of 0°) and Block 2 (constant AV discrepancy of 10°). VE, VAEi, and VAEc were each calculated separately for A and V responses. Correlations between VE, VAEi, and VAEc effects as well as correlations between effects and logMAR visual acuity were calculated as Spearman's rank correlations.

The size of the auditory and visual VE, VAEi, and VAEc was compared between groups by means of Welch ANOVAs and Games-Howell post-hoc comparisons. In addition, one-tailed, one-sample t tests were used to test whether effects were significantly larger than zero (Holm-corrected for multiple comparisons). All tests were additionally performed as Bayesian hypothesis tests using standard priors in JASP Version 0.14 (Wagenmakers et al., 2018) and Bayes Factors (BF₁₀ for two-tailed tests and BF₊₀/BF₋₀ for one-tailed tests) are reported to indicate the evidential value for the null or alternative hypothesis, respectively.

Model descriptions

We conducted modeling analysis to probe the computational components of the multisensory integration and learning processes in CC, DC, and SC individuals. Audiovisual integration is well described by Bayesian hierarchical causal inference models (Kayser and Shams, 2015; Körding et al., 2007; Wozny and Shams, 2011). The general idea is that cue integration improves perceptual precision and accuracy only when the cues provide redundant information about the same object or event, otherwise the cues should be processed separately. Therefore, the perceptual system needs to perform causal inference to determine which cues originated from the same cause and should be integrated. The causal inference (CI) models estimate how likely the auditory and visual cues share a common cause and arbitrate between integrating and segregating the cues accordingly. Not only do these models provide normative benchmarks against which behavioral VE data can be compared, they also allow quantitative assessment of any changes associated with the VAE, which are reflected in the changes of model parameters (Wozny and Shams, 2011).

We considered three different CI models, as well as four baseline models, for the audiovisual localization responses of each participant. All the models assume that the auditory (A) and visual (V) stimuli at the locations (s_A, s_V) give rise to noisy internal sensations (x_A, x_V), and that (x_A, x_V) vary from trial to trial even for the same (s_A, s_V). The models have access only to the noisy (x_A, x_V) and must estimate (s_A, s_V). Below we describe the important components of the CI models and the baseline models.

The Causal Inference (CI) Models. The CI models infer the unknown causal structure (C) of the stimuli in order to estimate the true locations (s_A, s_V). A and V may have a common cause (C = 1) or independent causes (C = 2). If A and V are caused by the same object (C = 1), then the optimal approach to estimate

the one true location ($s = s_A = s_V$) is to integrate the model's A and V location estimates (\hat{S}_A, \hat{S}_V) according to the relative reliabilities of the A and V sensory cues. This approach is often referred to as optimal integration (Alais and Burr, 2004; Rohde et al., 2016) or forced fusion (De Winkel et al., 2015). If A and V are caused by independent objects ($C = 2$), then the true A and V locations should be separately estimated, based solely on each unimodal sensory cue. The CI models are probabilistic models that represent information uncertainty in terms of probabilistic distributions (Li et al., 2020). Each CI model takes the perspective of a Bayesian ideal observer (Geisler, 2011), which combines sensory evidence and prior belief to calculate the probability of the underlying causal structure C, according to the Bayes' rule:

$$p(C|x_A, x_V) = \frac{p(x_A, x_V|C)p(C)}{p(x_A, x_V)} \quad (\text{Equation 1})$$

Here, C could be 1 (common cause) or 2 (independent causes), and $p(C|x_A, x_V)$ is the posterior probability ("posterior") of C given the sensory data (x_A, x_V). The prior probability ("prior") for the causal structure is denoted as $p(C)$. Specifically, the prior for a common cause, $p(C = 1)$, is known as the *causal prior* (Körding et al., 2007) and hereafter denoted as P_C for short. P_C represents the prior belief about how likely the cues share a common cause. Because the models only consider two possible causal structures, the prior for independent causes, $p(C = 2)$, simply equals $1 - P_C$. The likelihood function ("likelihood"), $p(x_A, x_V|C)$, calculates the joint probability of the sensory data (x_A, x_V) given a certain causal structure C. The likelihood quantifies the probabilistic mapping between the physical environment and the internal sensory representations.

For a common cause ($C = 1$), the posterior probability in Equation (1) is

$$p(C = 1|x_A, x_V) = \frac{p(x_A, x_V|C = 1)P_C}{p(x_A, x_V|C = 1)P_C + p(x_A, x_V|C = 2)(1 - P_C)} \quad (\text{Equation 2})$$

Like the priors, the posteriors sum to one; therefore, the posterior for $C = 2$, $p(C = 2|x_A, x_V)$, simply equals $1 - p(C = 1|x_A, x_V)$. The joint likelihood of getting the sensory data (x_A, x_V) under $C = 1$ is calculated by integration:

$$p(x_A, x_V|C = 1) = \int p(x_A, x_V|s)p(s)ds = \int p(x_A|s)p(x_V|s)p(s)ds \quad (\text{Equation 3})$$

where $s = s_A = s_V$. The *location prior*, $p(s)$, is the prior probability distribution over the possible stimulus locations; it represents the expectation of where the stimulus is likely to occur. Given a stimulus at location s , $p(x_A|s)$ and $p(x_V|s)$ are the unimodal likelihoods of sensing x_A and x_V , respectively. Assuming Gaussian distributions for $p(s)$, $p(x_A|s)$, and $p(x_V|s)$, Equation (3) can be analytically solved as

$$p(x_A, x_V|C = 1) = \frac{1}{2\pi\sqrt{\sigma_A^2\sigma_V^2 + \sigma_A^2\sigma_P^2 + \sigma_V^2\sigma_P^2}} \exp\left[-\frac{1}{2} \frac{(x_V - x_A)^2\sigma_P^2 + (x_A - \mu_P)^2\sigma_V^2 + (x_V - \mu_P)^2\sigma_A^2}{\sigma_A^2\sigma_V^2 + \sigma_A^2\sigma_P^2 + \sigma_V^2\sigma_P^2}\right] \quad (\text{Equation 4})$$

Here, μ_P and σ_P are the mean and standard deviation of the Gaussian location prior $p(s)$, and σ_A and σ_V are the standard deviations of the Gaussian noises that have corrupted the unisensory A and V signals, respectively.

By contrast, if the cues have independent causes ($C = 2$), then the joint likelihood of observing the sensory data (x_A, x_V) is calculated by

$$p(x_A, x_V|C = 2) = \iint p(x_A, x_V|s_A, s_V)p(s_A, s_V)ds_A ds_V = \left(\int p(x_A|s_A)p(s_A)ds_A\right) \left(\int p(x_V|s_V)p(s_V)ds_V\right) \quad (\text{Equation 5})$$

where s_A and s_V independently give rise to x_A and x_V . Again, assuming Gaussian distributions for $p(s_A)$, $p(s_V)$, $p(x_A|s_A)$, and $p(x_V|s_V)$, Equation 5 has the analytical solution

$$p(x_A, x_V|C = 2) = \frac{1}{2\pi\sqrt{(\sigma_A^2 + \sigma_P^2)(\sigma_V^2 + \sigma_P^2)}} \exp\left[-\frac{1}{2} \left(\frac{(x_A - \mu_P)^2}{\sigma_A^2 + \sigma_P^2} + \frac{(x_V - \mu_P)^2}{\sigma_V^2 + \sigma_P^2}\right)\right] \quad (\text{Equation 6})$$

A common method to obtain optimal estimation is to minimize the mean squared error as the cost function (Körding et al., 2007):

$$\text{Cost} = (\hat{S}_A - S_A)^2 + (\hat{S}_V - S_V)^2 \quad (\text{Equation 7})$$

Minimizing this cost function under $C = 1$ will give the analytical solution

$$\hat{S}_{A, C=1} = \hat{S}_{V, C=1} = \frac{\frac{X_A}{\sigma_A^2} + \frac{X_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}} \quad (\text{Equation 8})$$

This is the optimal estimate if $C = 1$. Equation (8) is a more general form of the optimal integration model (Alais and Burr, 2004; Ernst and Banks, 2002). If the models do not account for the location prior (which is equivalent to using a uniform location prior: $\mu_P = 0, \sigma_P \rightarrow \infty$), then Equation (8) reduces to the optimal integration model, which weights the sensory data by their relative reliabilities, and the reliability is represented by the inverse variance $1/\sigma_A^2$ or $1/\sigma_V^2$.

Under $C = 2$, minimizing the cost function in Equation (7) will give the analytical solutions for the A and V estimates separately as:

$$\hat{S}_{A, C=2} = \frac{\frac{X_A}{\sigma_A^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}} \quad (\text{Equation 9a})$$

$$\hat{S}_{V, C=2} = \frac{\frac{X_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}} \quad (\text{Equation 9b})$$

These are the optimal estimates if $C = 2$, which are equivalent to estimating the A and V locations independently, based solely on the respective unisensory data.

The three CI models we considered are identical in applying Equations (1, 2, 3, 4, 5, 6, 7, 8, 9a, and 9b) but differ in how they combine the estimates obtained under $C = 1$ (integration) and $C = 2$ (segregation) to generate the responses for A and V localization. The optimal strategy to combine the estimates is model averaging (MA), because it results in the minimal mean expected squared error (Körding et al., 2007). The CI model with the MA strategy (CI-MA) calculates the weighted averages of the estimates under $C = 1$ and $C = 2$; the corresponding weights are the posterior probabilities of $C = 1$ and $C = 2$:

$$\hat{S}_A = p(C = 1|X_A, X_V)\hat{S}_{A, C=1} + (1 - p(C = 1|X_A, X_V))\hat{S}_{A, C=2} \quad (\text{Equation 10a})$$

$$\hat{S}_V = p(C = 1|X_A, X_V)\hat{S}_{V, C=1} + (1 - p(C = 1|X_A, X_V))\hat{S}_{V, C=2} \quad (\text{Equation 10b})$$

In addition to the MA strategy, behavioral studies on causal inference have shown that humans and non-human primates sometimes apply two heuristic alternatives for audiovisual localization (Mohr et al., 2020; Wozny et al., 2010): probability matching (PM) and model selection (MS). The CI model with the PM strategy (CI-PM) chooses between the estimates under $C = 1$ and $C = 2$ with the posterior probabilities of $C = 1$ and $C = 2$:

$$\hat{S}_A = \begin{cases} \hat{S}_{A, C=1} & \text{if } p(C = 1|X_A, X_V) > \xi \\ \hat{S}_{A, C=2} & \text{if } p(C = 1|X_A, X_V) \leq \xi \end{cases} \quad (\text{Equation 11a})$$

$$\hat{S}_V = \begin{cases} \hat{S}_{V, C=1} & \text{if } p(C = 1|X_A, X_V) > \xi \\ \hat{S}_{V, C=2} & \text{if } p(C = 1|X_A, X_V) \leq \xi \end{cases} \quad (\text{Equation 11b})$$

where ξ is sampled from a uniform distribution $[0, 1]$ on each trial. For example, if the posterior probability of $C = 1$ is 0.6 on a trial, then 60% of the time the CI-PM model will choose $\hat{S}_{A, C=1}$ as its final estimate for the A location, but 40% of the time it will choose $\hat{S}_{A, C=2}$ as its final estimate for the A location (Wozny et al., 2010). The CI model with the MS strategy (CI-MS) compares the posterior probabilities of $C = 1$ and $C = 2$ and simply uses the estimates from the more probable causal structure, that is, the causal structure whose posterior probability is >0.5 :

$$\hat{S}_A = \begin{cases} \hat{S}_{A, C=1} & \text{if } p(C = 1|X_A, X_V) > .5 \\ \hat{S}_{A, C=2} & \text{if } p(C = 1|X_A, X_V) \leq .5 \end{cases} \quad (\text{Equation 12a})$$

$$\hat{S}_V = \begin{cases} \hat{S}_{V, C=1} & \text{if } p(C=1|X_A, X_V) > .5 \\ \hat{S}_{V, C=2} & \text{if } p(C=1|X_A, X_V) \leq .5 \end{cases} \quad (\text{Equation 12b})$$

For example, if the posterior probability of $C=1$ is 0.6 on a trial, then the CI-MS model will choose $\hat{S}_{A,C=1}$ as its final estimate for the A location (Wozny et al., 2010).

Baseline Models. In addition to the three causal inference models (CI-MA, CI-PM, and CI-MS), we considered four baseline models. The baseline models do not perform causal inference; instead, they assume a fixed causal structure and always either fuse or segregate the A and V cues to estimate the stimulus locations.

Two of the baseline models take into account the location prior: (1) A forced fusion (FF) model with a Gaussian location prior. This model always integrates A and V sensory data based on their reliabilities according to Equation (8). The FF model assumes a common cause and is equivalent to the $C=1$ branch of a CI model. (2) A segregation (SG) model with a Gaussian location prior. This model completely segregates A and V sensory data without any integration or inter-modal influence. The A and V locations are estimated independently, based solely on the information within each modality according to Equations (9a) and (9b). The SG model assumes independent causes and is equivalent to the $C=2$ branch of a CI model.

To check whether the prior parameters were necessary for modeling our data, we additionally tested the simplified versions of the above two baseline models without the location prior: (3) A forced fusion model with no priors (FFnp), and (4) a segregation model with no priors (SGnp). Mathematically, having no priors is equivalent to assuming a flat location prior ($\mu_P=0$, $\sigma_P \rightarrow \infty$), which means these models have no expectations for the stimulus location, and a fixed causal prior ($P_C=1$ for FFnp, $P_C=0$ for SGnp). The FFnp model is identical to the optimal integration model (Alais and Burr, 2004; Ernst and Banks, 2002) and applies a simplified version of Equation (8) without the prior parameters:

$$\hat{S}_{A, C=1} = \hat{S}_{V, C=1} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} \quad (\text{Equation 13})$$

This equation is the same as the reliability-based weighted-average formula of the optimal integration model. Similarly, the SGnp model applies a simplified version of Equations (9a) and (9b) without the prior parameters:

$$\hat{S}_{A, C=2} = x_A; \hat{S}_{V, C=2} = x_V \quad (\text{Equation 14})$$

Model Parameters and Simulations

Causal inference models. Each CI model has seven parameters: Δ_A, σ_A —the bias and the standard deviation (SD) of the auditory Gaussian likelihood; Δ_V, σ_V —the bias and the SD of the visual Gaussian likelihood; μ_P, σ_P —the mean and the SD of the Gaussian location prior for stimuli of either modality; P_C —the causal prior, that is, the prior for a common cause.

Typical CI models assume unbiased likelihoods and priors (e.g., Körding et al., 2007; Wozny et al., 2010); that is, they assume that the likelihoods center at the true stimulus locations ($\Delta_A=0$, $\Delta_V=0$) and that the location prior centers at 0° azimuth angle ($\mu_P=0$). Here we incorporate non-zero bias terms to probe whether the VAEc is associated with changes in the likelihood or prior biases – in other words, whether a shift in the likelihood or location prior distribution is a possible computational mechanism underlying the VAEc (for detail, see Wozny and Shams, 2011).

For each combination of the seven parameters and each combination of true audiovisual (AV) stimulus locations (s_A, s_V), each CI model performed 10,000 Monte Carlo simulated AV localization trials. The noisy internal sensory representations x_A and x_V were generated according to the Gaussian likelihood functions $p(x_A|s)$ and $p(x_V|s)$: On each simulated trial, x_A and x_V were independently sampled from two Gaussian distributions representing the independent Gaussian noises that corrupted the A and V signals. The Gaussian distributions centered at the true stimulus locations s_A and s_V plus the likelihood biases Δ_A and Δ_V , respectively, with σ_A and σ_V determining the strengths of the Gaussian noises: $x_A \sim \mathcal{N}(s_A + \Delta_A, \sigma_A)$ and $x_V \sim \mathcal{N}(s_V + \Delta_V, \sigma_V)$. Additionally, the model accounted for a prior bias for the stimulus location,

represented by a Gaussian distribution centered at μ_p , and the strength of this location prior bias was determined by σ_p : $p(s) = \mathcal{N}(\mu_p, \sigma_p)$. The model combined the likelihoods, the location prior, and the causal prior to calculate the posterior posteriors for $C = 1$ and $C = 2$, as well as estimating the A and V locations under each causal structure. Up to this step, the three CI models would produce the same estimates given a certain set of parameter values. The CI models differed in the next step: Each CI model implemented its decision strategy (MA, PM, or MS) to use these estimates to generate the responses for the A and V locations on that AV trial. The resulting 10,000 simulated A responses and 10,000 simulated V responses were then each binned into a histogram of 12 bins, corresponding to the 12 discrete button-response positions. The trial counts in the bins were then normalized into relative proportions of counts. The final output of each CI model were two histograms of proportions of counts, one for A localization and one for V localization. These were the model's predicted AV response distributions (in proportions of counts) for a pair of AV stimulus locations (s_A, s_V), given a set of parameter values.

Baseline Models. The two baseline models with a Gaussian location prior – the forced fusion (FF) model and the segregation (SG) model – have the same parameters as the CI models except that the causal prior (P_c) is fixed instead of being a free parameter. The FF model assumes a common cause, so $P_c = 1$. The SG model assumes independent causes, so $P_c = 0$. The simulation processes of the FF model and the SG model are identical to the $C = 1$ and $C = 2$ branches, respectively, of the CI models.

The other two baseline models – forced fusion with no priors (FFnp) and segregation with no priors (SGnp) – do not have any location prior parameters. For the purpose of calculating average parameter values, the causal prior is assigned to be $P_c = 1$ for FFnp and $P_c = 0$ for SGnp. In other words, FFnp and SGnp each has only four parameters: Δ_A, σ_A (mean bias and SD of the A likelihood) and Δ_V, σ_V (mean bias and SD of the V likelihood).

Like the CI models, each baseline model performed 10,000 Monte Carlo simulated AV localization trials and generated model-predicted AV response distributions (in proportions of counts). For the two forced fusion models (FF and FFnp), the predicted A and V response distributions were identical, because the models assumed that A and V shared the same location on each trial. For the two segregation models (SG and SGnp), the predicted A and V response distributions were different.

Model fitting and comparison

Step 1. The unimodal A and V Gaussian likelihood biases and SDs ($\Delta_A, \sigma_A, \Delta_V, \sigma_V$) were estimated based on the unimodal A and V trials of Blocks 1 and 2 separately for each participant (Zierul et al., 2019). To fit the unimodal A and V localization data, we applied the “fitdist” function from the R-package “fitdistrplus” (package version 1.1-3, R version 4.0.4), using maximum likelihood estimation (MLE) and discretizing the Gaussian distributions to match the discrete experimental data. For each participant, we obtained a set of best-fitted ($\Delta_A, \sigma_A, \Delta_V, \sigma_V$) for Block 1 and another set for Block 2.

Step 2. We fitted and compared three CI models and four baseline models to each participant's AV trials in Block 1, to find the individual's best-fitting model and corresponding prior parameters (e.g., μ_p, σ_p , and P_c for any CI model). This step identified the possible computational components that best described the observed multisensory integration patterns (as indicated by the VE). For a given participant, all seven models used the same individually fitted unimodal parameters ($\Delta_A, \sigma_A, \Delta_V, \sigma_V$) for Block 1 from Step 1. Unlike the simple Gaussian likelihoods in Step 1, the model-predicted responses in this step were complex distributions, hence we customized the fitting procedure. Again, we used the MLE method to find the model and parameters that best explained the data. For the discrete button-response outcomes, the likelihood function to be maximized was the multinomial log-likelihood (Cao et al., 2019; Körding et al., 2007):

$$LL = \sum_i^k n_i \ln(p_i) + \ln C \quad (\text{Equation 15})$$

LL denotes the multinomial log-likelihood, $k = 12$ is the number of button-response positions, n_i denotes the observed counts of experimental trials at button-response position i , p_i denotes the model-predicted response probability at button-response position i , and C is the multinomial coefficient determined by the observed counts. $\ln C$ is a constant for all models and parameter sets, so maximizing LL is equivalent to maximizing $\sum_i^k n_i \ln(p_i)$. For each set of prior parameters of each prior-including model (i.e., the three CI

models, the FF model, and the SG model), and for each pair of A and V stimulus locations (s_A, s_V), 10,000 simulated AV trials were run to generate 10,000 A responses and 10,000 V responses in relative proportions of counts at the button-response positions (see “Model Parameters and Simulations” above); these model-predicted proportions were entered into Equation (15) to approximate the response probabilities p_i and calculate the log-likelihood for (s_A, s_V). The log-likelihoods for all pairs of (s_A, s_V) were then summed to get the overall log-likelihood. The best-fitting model and prior parameters were the ones that maximized this overall log-likelihood, or equivalently, minimized its negative value (Cao et al., 2019; Körding et al., 2007). To optimize the fitting procedure, we applied the “mle” function from the R-package “stats4” (version 4.0.4) and implemented a bounded search method (L-BFGS-B) to minimize the negative log-likelihood (Byrd et al., 1995). The search was initialized using the best starting prior parameter values obtained from a uniform grid search (Mohl et al., 2020), which divided each free prior parameter space into 10 even portions (i.e., 11 values). The grid yielded $11^3 = 1,331$ initial prior parameter sets for each CI model, and $11^2 = 121$ initial prior parameter sets for the FF and the SG models.

To compare the model performances, we calculated the Bayesian Information Criterion (BIC) for each model: $BIC = -2(\widehat{LL}) + k * \ln(N)$, where (\widehat{LL}) denotes the maximum log-likelihood, k is the number of parameters, and N is the total number of data points (Neath and Cavanaugh, 2012). The BIC awards models that fit the data well but penalizes models with more parameters (i.e., more complex). For a given dataset, the best model is the one with the lowest BIC. We thus identified the best model for each participant. All participants were best-fitted by models with priors, that is, no participant was best-fitted by FFnp or SGnp.

Step 3. To probe whether crossmodal learning (as indicated by VAEc) was associated with changes in any likelihood or prior parameters, we conducted modeling analysis on each participant’s Block 2 data and compared the results with those from Block 1. The A and V likelihood parameters ($\Delta_A, \sigma_A, \Delta_V, \sigma_V$) for Block 2 were already fitted individually in Step 1. We used these best-fitting likelihood parameters from Step 1, as well as the best model determined in Step 2, to fit each participant’s Block 2 AV trials and obtain the best-fitting prior parameters for Block 2. We assumed that the individual’s best model did not change from Block 1 to Block 2, but that the model parameters could change (Wozny and Shams, 2011). The model fitting procedure was identical to that in Step 2. There was no model comparison in this step because the same best models from Step 2 were applied.