PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception

Fangfang Hong[1]*, **Stephanie Badde**[2☯], **Michael S. Landy**[1,3☯]

1 Department of Psychology, New York University, New York City, New York, United States of America, **2** Department of Psychology, Tufts University, Medford, Massachusetts, United States of America, **3** Center for Neural Science, New York University, New York City, New York, United States of America

☯ These authors contributed equally to this work.
* fh862@nyu.edu

Citation: Hong F, Badde S, Landy MS (2021) Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. PLoS Comput Biol 17(11): e1008877. https://doi.org/10.1371/journal.pcbi.1008877

Editor: Ulrik R. Beierholm, Durham University, UNITED KINGDOM

Received: February 26, 2021

Accepted: October 26, 2021

Published: November 15, 2021

Data Availability Statement: All data and code files are available on OSF database (https://osf.io/6mt7x/).

Funding: This work was supported by NIH grant EY08266, awarded to MSL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

To obtain a coherent perception of the world, our senses need to be in alignment. When we encounter misaligned cues from two sensory modalities, the brain must infer which cue is faulty and recalibrate the corresponding sense. We examined whether and how the brain uses cue reliability to identify the miscalibrated sense by measuring the audiovisual ventriloquism aftereffect for stimuli of varying visual reliability. To adjust for modality-specific biases, visual stimulus locations were chosen based on perceived alignment with auditory stimulus locations for each participant. During an audiovisual recalibration phase, participants were presented with bimodal stimuli with a fixed perceptual spatial discrepancy; they localized one modality, cued after stimulus presentation. Unimodal auditory and visual localization was measured before and after the audiovisual recalibration phase. We compared participants' behavior to the predictions of three models of recalibration: (a) Reliability-based: each modality is recalibrated based on its relative reliability—less reliable cues are recalibrated more; (b) Fixed-ratio: the degree of recalibration for each modality is fixed; (c) Causal-inference: recalibration is directly determined by the discrepancy between a cue and its estimate, which in turn depends on the reliability of both cues, and inference about how likely the two cues derive from a common source. Vision was hardly recalibrated by audition. Auditory recalibration by vision changed idiosyncratically as visual reliability decreased: the extent of auditory recalibration either decreased monotonically, peaked at medium visual reliability, or increased monotonically. The latter two patterns cannot be explained by either the reliability-based or fixed-ratio models. Only the causal-inference model of recalibration captures the idiosyncratic influences of cue reliability on recalibration. We conclude that cue reliability, causal inference, and modality-specific biases guide cross-modal recalibration indirectly by determining the perception of audiovisual stimuli.

Author summary

Audiovisual recalibration of spatial perception occurs when we receive audiovisual stimuli with a systematic spatial discrepancy. The brain must determine to which extent both modalities should be recalibrated. In this study, we scrutinized the mechanisms the brain employs to do so. To this aim, we conducted a classical audiovisual recalibration experiment in which participants were adapted to spatially discrepant audiovisual stimuli. The visual component of the bimodal stimulus was either less, equally, or more reliable than the auditory component. We measured the amount of recalibration by computing the difference between participants' unimodal localization responses before and after the audiovisual recalibration. Across participants, the influence of visual reliability on auditory recalibration varied fundamentally. We compared three models of recalibration. Only a causal-inference model of recalibration captured the diverse influences of cue reliability on recalibration found in our study, this model is also able to replicate contradictory results found in previous studies. In this model, recalibration depends on the discrepancy between a sensory measurement and the perceptual estimate for the same sensory modality. Cue reliability, perceptual biases, and the degree to which participants infer that the two cues come from a common source govern audiovisual perception and therefore audiovisual recalibration.

Introduction

Cross-modal integration

In our daily lives, we continuously estimate properties of the environment such as the location of a barking dog. Usually multiple sensory cues for each property arrive in the brain, a glimpse of the dog's wagging tail and the barking both give away its location. However, due to external noise in the environment and internal noise in our sensory systems, two cues hardly ever agree perfectly. To still form a coherent percept, the brain relies on a weighted mixture of the cues. This strategy becomes evident when the cues are in conflict. For example, when a ventriloquist speaks without moving her lips, the auditory signal indicates that the speech originates from the ventriloquist, while the visual signal indicates the "dummy". Typically, vision dominates the combined spatial estimate of the sound source. In this example, we perceive the "dummy" to be speaking. This phenomenon is called the ventriloquism effect [1–3]. However, when visual reliability, the inverse of the average variability of a cue, is degraded enough to be lower than auditory reliability, the combined spatial estimate is no longer dominated by the visual but instead by the auditory cue, indicating that cue integration depends on their relative reliability [4–6]. This integration strategy maximizes the precision of the estimate, i.e., reduces its variability. In addition to audiovisual spatial perception, reliability-based cue integration has also been found for visual and auditory cues to temporal rate [7–9], visual and haptic size and shape [10–12] as well as numerosity [13], and visual and vestibular cues to heading direction [14–19].

If during a ventriloquist's show, the speech sounds originate from someone standing behind the stage or if the dummy's mouth moves out of synch with the speech sounds, the audience is unlikely to experience a ventriloquism effect. In other words, integration breaks down when the two cues are too different to be perceived as coming from a common source. Such a breakdown of integration with spatial and temporal cue conflicts has been found not only in auditory-visual spatial integration [20–25], but also in integration of visual and

vestibular heading-direction signals [26, 27] and integration of visual and haptic surface thickness [28]. These findings suggest that the brain infers the causal relationship underlying cues from multiple modalities. A Bayesian observer would not simply decide between integrating the cues or keeping them segregated, but rather combine estimates based on integration and segregation, each weighted by the probability of the underlying causal scenario [29]. This causal-inference model accurately captures the integration of audiovisual spatial signals [29] and has been used to explain cross-modal perception in various contexts [26, 27, 30–40].

Cross-modal recalibration

Reliability-weighted sensory cue integration maximizes the precision of the final estimate but not its accuracy, the agreement between the cue with the property of interest. Estimates are dominated by the more precise not by the more accurate cue. This preference of precision over accuracy might be rooted in the fact that single sensory cues can contain information about their reliability [41], but information about accuracy is impossible to derive from a cue on its own. Only systematic disparities between two sensory cues indicate a problem with their accuracy and, at the same time, provide a chance to resolve this issue. In the case of mismatched cues, the brain should adapt the interpretation of the cues to reduce that disparity and so, hopefully, maximize accuracy. Indeed, after repeated exposure to spatially discrepant audiovisual stimuli, the shift in auditory spatial perception toward the visual stimulus is still present when the visual signal is no longer available. This persisting shift is called the ventriloquism aftereffect, and has been replicated across different modalities [32, 42–51].

In many spatial-aftereffect studies, vision serves as the "teaching signal" that is used to calibrate auditory or tactile perception. Some studies found a discrepant auditory stimulus presented during adaptation can induce systematic shifts in visual localization, but the aftereffects were not as robust as the recalibration of auditory spatial perception [47, 52, 53]. This dominance of vision seems reasonable given that visual information is usually more accurate and reliable than information from other modalities in terms of localizing objects spatially [54].

Reliability-based cross-modal recalibration. Which modality serves as the "teaching signal" in cross-modal recalibration might be determined based on cue reliability. Consistent with this hypothesis, more robust evidence for recalibration of visual perception was found in studies that examined aftereffects for properties for which the visual signal was not as reliable as the other signal [55–60]. Burge and colleagues directly tested the relationship between cue reliability and the amount of recalibration by manipulating the reliability of a visual stimulus for slant to be either smaller, almost equal, or greater than that of a haptic cue to slant [61]. As the visual cue became less reliable, greater recalibration of vision and less recalibration of haptics was observed. The authors concluded, in agreement with others [53, 59], that each sense is recalibrated proportional to its relative reliability.

Fixed-ratio cross-modal recalibration. In contradiction to the results just discussed, a recent study investigating the mechanism underlying visual-vestibular recalibration found evidence that the two modalities are recalibrated in a fixed ratio regardless of cue reliability [62]. More specifically, after exposing humans and monkeys to systematically discrepant visual and vestibular heading directions, both cues significantly shifted in the direction required to reduce cue conflict, but the amount of recalibration in either modality did not depend on the measured relative cue reliabilities [62]. A model assuming a fixed ratio between the degree of recalibration of each sense, independent of relative reliability, best captured the data.

Causal inference and cross-modal recalibration. Both reliability-based and fixed-ratio models of cross-modal recalibration assume that the brain acts upon the discrepancy between the two sensory cues. However, according to the principles of cross-modal integration, two

discrepant cues can lead to very different percepts, depending on each cue's reliability and inference about whether they have common or separate origins. Cross-modal recalibration might take this perceptual inference into account. Indeed, causal-inference models of cross-modal recalibration have successfully predicted visual-auditory [63] and visual-tactile [32] ventriloquism aftereffects. In these models, recalibration is not based on a mere comparison of two sensory cues, but rather relates the cues to the perceptual estimates and by doing so incorporates causal inference and cue reliability. According to this model, conflicting findings regarding the influence of cue reliability on recalibration might reflect differences in the perceptual estimates rather than diverging underlying mechanisms.

Preview

In this study, we contrasted all three accounts of cross-modal recalibration by fitting the models to data from an audiovisual ventriloquism-aftereffect study. Across sessions, we manipulated cue reliability, a determinant of reliability-based and causal-inference-driven cross-modal recalibration. Additionally, we controlled for the effect of modality-specific biases on the spatial perception of the two sensory cues by choosing visual stimulus locations that matched perceptually with the locations of the auditory stimuli for each participant.

With decreased visual reliability, many participants showed either increasing or nonlinearly changing auditory recalibration. No clear pattern of visual recalibration was found. These results cannot be explained by either the reliability-based or the fixed-ratio model. The causal-inference model, on the other hand, is able to capture these idiosyncratic effects based on individual differences in cue reliability, modality-specific biases, and an a priori belief about how often visual and auditory cues come from a common source. Thus, the model comparison suggests that cross-modal recalibration is driven by a comparison between sensory cues and perceptual estimates, which in turn are determined using causal inference.

Results

Cue reliability

In the first part of the study, spatial reliability for one auditory stimulus and three visual stimuli was estimated for each participant using a unimodal spatial-discrimination task (Fig 1B; see Materials and methods). Visual reliability had been manipulated by varying the horizontal spread of a random collection of ten Gaussian blobs (Fig 1A). For each stimulus condition, we fitted a cumulative Gaussian distribution to the responses as a function of test stimulus location (Fig 2A, mean adjusted $R^2 = 0.953$, range = 0.751–0.997). To compare spatial-discrimination performance across stimulus conditions, we computed the just-noticeable difference (JND; Fig 2B and 2C). Statistical analysis of the JNDs (Section S1.1 in S1 Appendix) confirmed that visual stimulus reliabilities were (i) smaller than, (ii) comparable to, and (iii) larger than the auditory reliability.

Modality-specific spatial biases

In the second part of the study, we measured participants' modality-specific biases in the spatial perception of auditory stimuli relative to a visual stimulus (Fig 1A) using a bimodal spatial-discrimination task (Fig 3A). We did so at four auditory stimulus locations and for visual stimuli with high spatial reliability. Four separate psychometric functions were fitted, one for each auditory stimulus location (Fig 3B, mean adjusted $R^2 = 0.909$, range = 0.751–0.981). From each psychometric function, we calculated the point of subjective equality (PSE) and then described the four PSEs as a linear function of the underlying auditory location (Fig 4A).

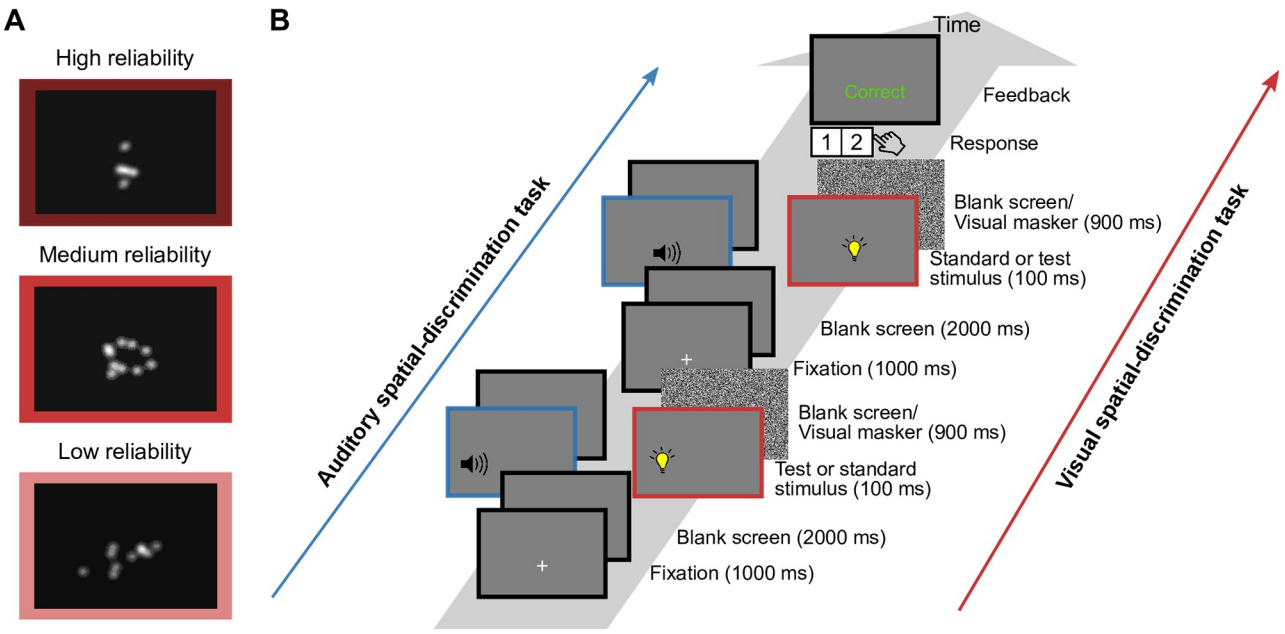

Fig 1. Visual stimuli and experimental procedure for the unimodal spatial-discrimination task. (A) Example visual stimuli (contrast exaggerated). (B) Task timing (blue: auditory stimuli; pink: visual stimuli). Participants were successively presented with a standard stimulus (located straight ahead) and a test stimulus (located to the left or right) in random order. After stimulus presentation, they used a keypad to report which interval contained the stimulus located farther to the right. Feedback was provided.

https://doi.org/10.1371/journal.pcbi.1008877.g001

The estimated slopes for five out of six participants were significantly larger than 1 (Fig 4B) indicating that the auditory stimuli were perceived as shifted towards the periphery relative to the visual stimuli. Four participants showed significant negative *y*-intercepts. That is, they perceived the auditory stimuli as shifted to the left relative to visual stimuli. From the linear regression line through the PSEs, we extracted the four visual stimulus locations that were perceived to be co-located with the four auditory stimulus locations and used these locations in all subsequent tasks.

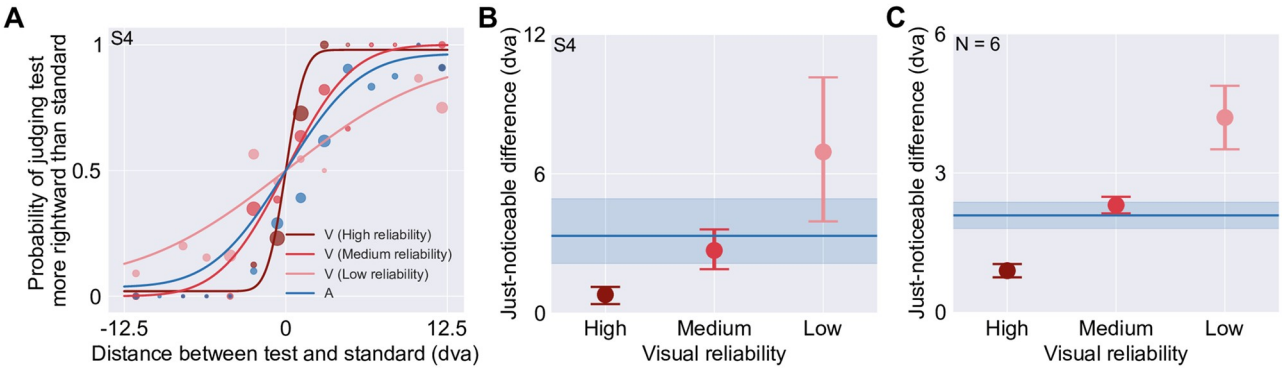

Fig 2. Results for the unimodal spatial-discrimination task. (A) Psychometric functions for representative participant S4. The probability of judging the test stimulus as to the right of the standard stimulus is plotted as a function of the distance between the two stimuli, with negative numbers indicating that the test stimulus was located to the left of the standard stimulus (presented straight ahead). V: Visual. A: Auditory. Filled circles: binned response proportions (bin size = 1.8˚). The area of each filled circle is proportional to the number of trials within the bin. Solid curves: psychometric functions fit to the data. (B) Estimated just-noticeable difference (JND) of each visual stimulus type for S4. Solid line: auditory JND. Error bars and blue area: 95% bootstrapped confidence intervals. (C) Group mean JNDs (± SEM).

https://doi.org/10.1371/journal.pcbi.1008877.g002

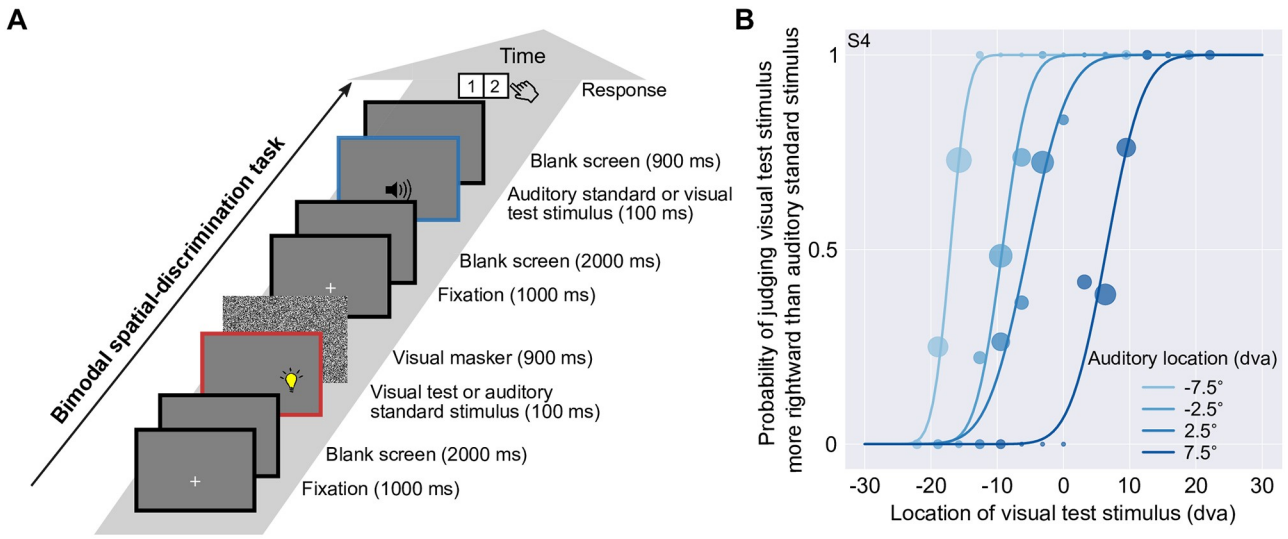

Fig 3. Experimental procedure and results for the bimodal spatial-discrimination task. (A) In each trial, a visual stimulus (high reliability) and an auditory stimulus (four possible auditory locations, ±2.5 and ±7.5˚ relative to straight-ahead) were presented in random order. Participants reported whether the visual stimulus was to the left or right of the auditory stimulus. Feedback was not provided. (B) Psychometric functions for participant S4. Probability of judging the visual to the right of the auditory stimulus is plotted as a function of visual stimulus location. Filled circles: binned response proportions (bin size = 3˚). Curves: psychometric functions fitted separately for the four auditory stimulus locations (shades of blue). The area of each filled circle is proportional to the number of trials in each bin.

https://doi.org/10.1371/journal.pcbi.1008877.g003

Localization response precision

In the next part of the study, we measured participants' localization noise unrelated to spatial perception (e.g., noise due to holding a location in memory and errors indicating the intended location). To this aim, participants performed a direct localization task with maximally reliable visual stimuli. At the same time, they were familiarized with our custom-made device used to

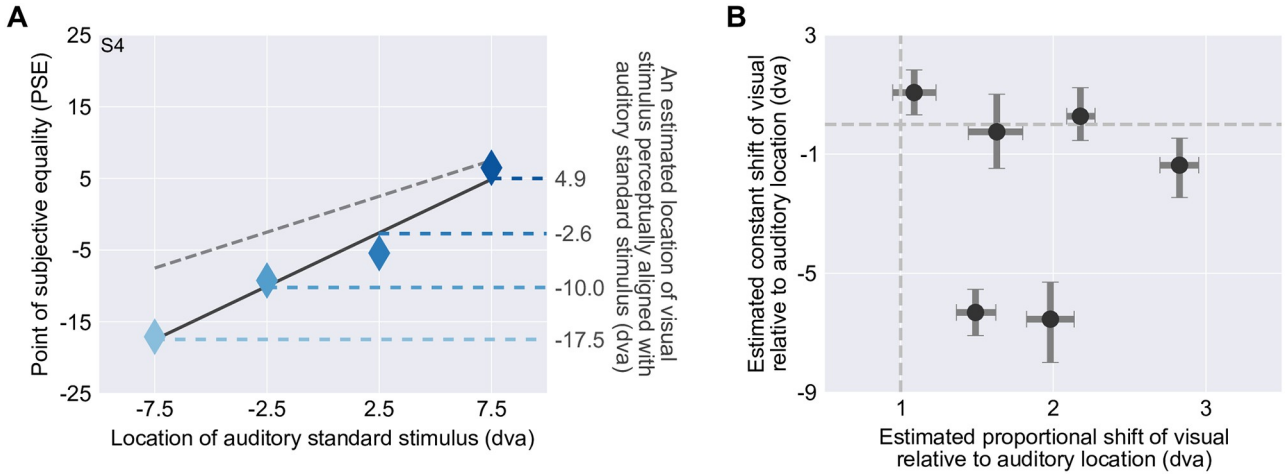

Fig 4. Modality-specific biases in spatial perception. (A) Point of subjective equality (PSE) as a function of the location of the auditory stimulus. Dashed grey line: identity line; solid line: linear regression line; horizontal dashed lines: visual stimulus locations perceived as co-located with the four auditory standard locations based on the regression. (B) Estimated constant, location-independent (*y*-axis; regression intercept) and proportional, location-dependent (*x*-axis; regression slope) shift of visual relative to auditory stimulus location. The proportional and constant shifts equal the slope and intercept of the linear regression line through the PSEs (see panel A). Error bars: 95% bootstrapped confidence intervals. Vertical and horizontal dashed lines correspond to the absence of proportional and constant shifts, respectively.

https://doi.org/10.1371/journal.pcbi.1008877.g004

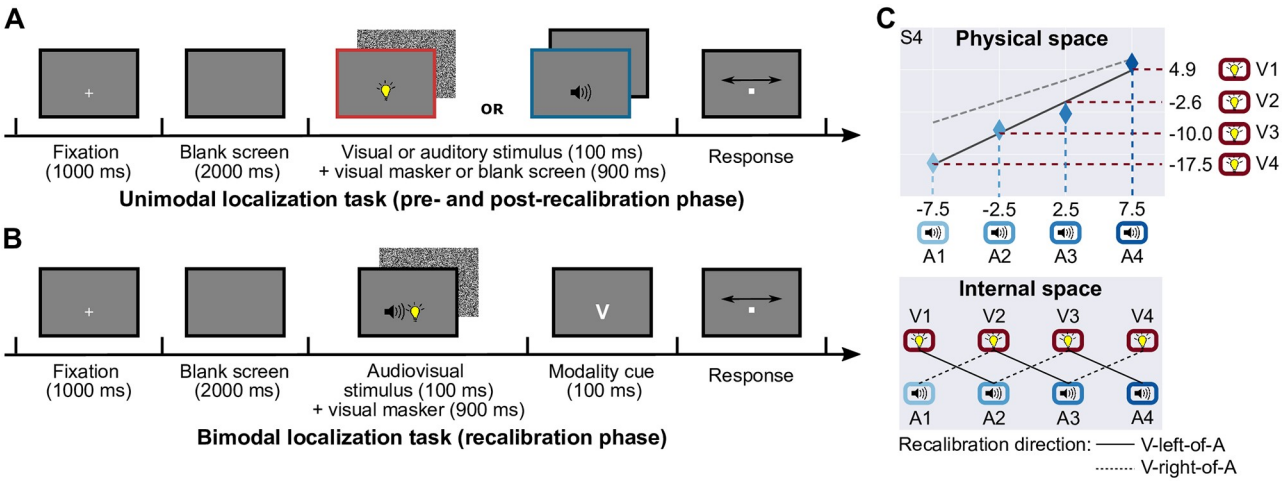

Fig 5. Experimental procedure and conditions during the recalibration experiment. (A) Timeline for unimodal localization tasks (pre- and post-recalibration phase). In each trial, either a visual or an auditory stimulus was presented; participants indicated its location using a visual cursor displayed on the screen. Feedback was not provided. (B) Timeline for the bimodal localization task (recalibration phase). In each trial, participants were presented with a spatially discrepant audiovisual stimulus pair; they were asked to localize one of the modalities, cued after stimulus presentation (V: localize the visual component, A: localize the auditory component). Feedback was not provided. (C) Stimulus locations in physical and perceptual space. Top panel: the physical locations of the four perceptually aligned audiovisual stimulus pairs identified at the beginning of the study for participant S4, stimuli were always presented at one of these locations; bottom row: pairs with a constant perceptual spatial discrepancy were presented during the recalibration phase, solid lines: location pairs presented in the visual-left-of-auditory condition; dashed lines: location pairs presented in the visual-right-of-auditory condition.

https://doi.org/10.1371/journal.pcbi.1008877.g005

move a visual cursor to the stimulus location. We assumed that localization errors were unbiased and independent of stimulus location and fitted them with a Gaussian distribution centered at zero to estimate the extent of noise corrupting localization responses (see Section S2 in S1 Appendix for a more complex model and results of a model comparison). Participants' spatial perception-unrelated localization noise was 1.85˚ on average (range 1.55–2.29˚).

Recalibration effects

In the final part of the study, participants completed six sessions of the audiovisual recalibration experiment (2 recalibration directions × 3 reliability levels). Each session consisted of three phases: (1) pre-recalibration: participants localized unimodally presented visual and auditory stimuli (Fig 5A); (2) recalibration: participants were presented with audiovisual stimulus pairs with a constant spatial discrepancy in perceptual space; they localized one modality cued after stimulus presentation (Fig 5B and 5C); (3) post-recalibration: the unimodal localization task was repeated.

To statistically examine recalibration of each modality, we fitted linear regressions separately to pre- and post-recalibration localization responses as a function of stimulus location (Fig 6A). Recalibration effects were computed as the difference between pre- and post-recalibration regression intercepts; shifts compensating for the audiovisual discrepancy present during the recalibration phase were coded as positive. As no significant effects of recalibration direction on the recalibration effect were found in our statistical analysis (Section S1.2 in S1 Appendix), we averaged the recalibration effect across the two recalibration directions (visual to right/left of auditory) for display (Fig 6B). For the majority of participants (three out of six), auditory recalibration was a non-monotonic function of visual stimulus reliability. Two participants showed a monotonic increase in auditory recalibration with decreasing visual stimulus reliability. One participant showed a monotonic decrease in auditory recalibration as visual

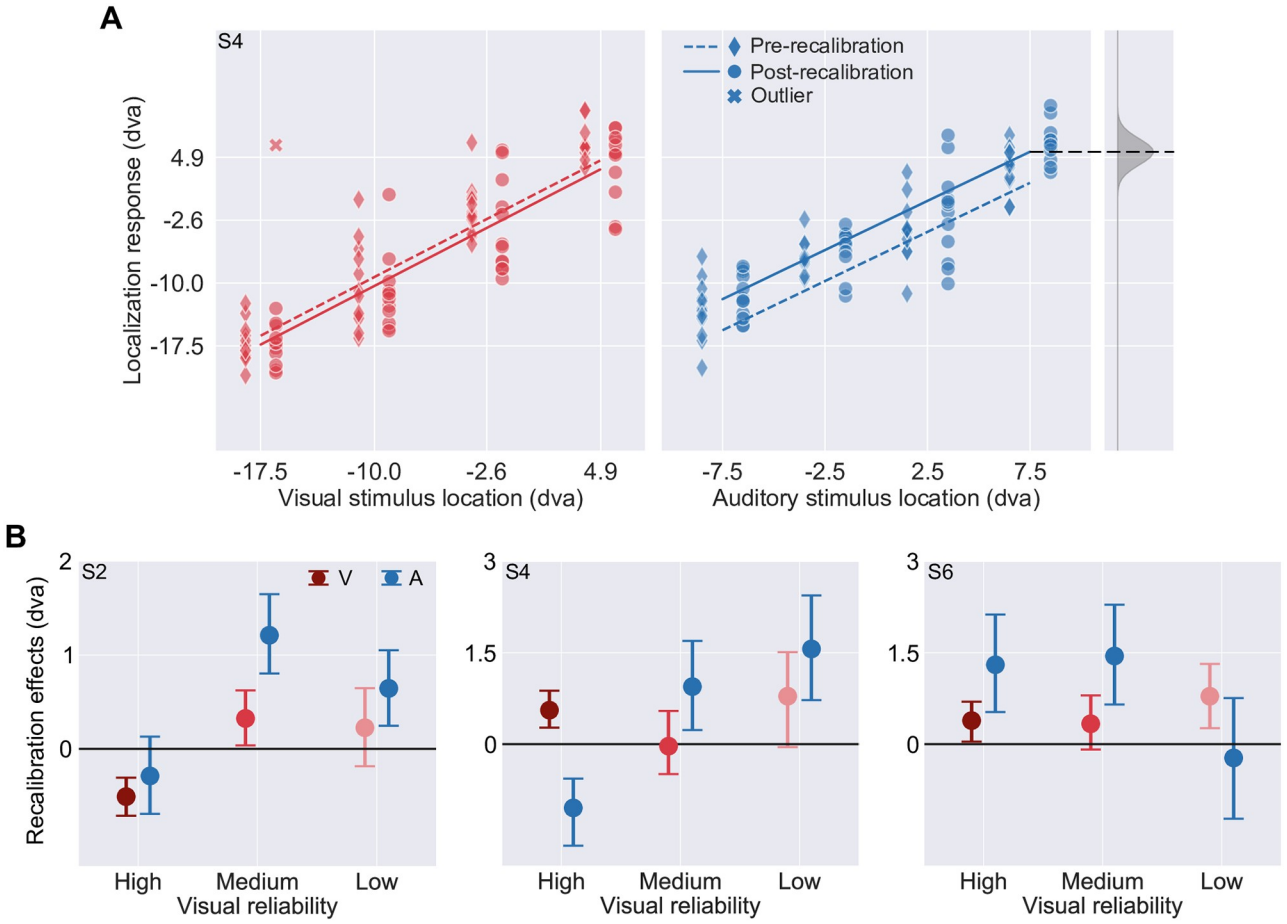

Fig 6. Recalibration effects. (A) Pre- (diamonds, jittered to the left) and post-recalibration (circles, jittered to the right) localization responses as a function of auditory and visual stimulus locations (relative to straight-ahead) measured for participant S4 in the visual-right-of-auditory and low-visual-reliability condition. Localization responses were summarized using linear regression (dashed lines: pre-recalibration; solid lines: post-recalibration). Grey shaded area: estimated probability distribution of spatial perception-unrelated localization noise centered at an example location. (B) Auditory and visual recalibration effects (the difference between the intercepts of the pre-and post-recalibration regression lines in panel A) as a function of visual stimulus reliability for three participants. Error bars: 95% bootstrapped confidence intervals.

https://doi.org/10.1371/journal.pcbi.1008877.g006

reliability decreased. Unsurprisingly, our statistical analysis of the recalibration effect revealed no significant main effects or interactions involving visual stimulus reliability (Section S1.2 in S1 Appendix).

Recalibration models

To understand the mechanisms of cross-modal recalibration, we compared participants' behavior to three existing models of cross-modal recalibration: (1) a reliability-based, (2) a fixed-ratio, and (3) a causal-inference model (for details see Models of audiovisual recalibration). All three models conceptualize recalibration as constant updating of a modality-specific shift applied to the measurements before the estimate is derived. However, these three models differ in their assumptions about the way in which the amount of recalibration for each modality is determined. As a consequence, these three models make different predictions for the influence of visual reliability on audiovisual recalibration.

The reliability-based model. According to this model, the brain determines the amount of recalibration (i.e., the measurement shift) for each modality based on the reliabilities of both

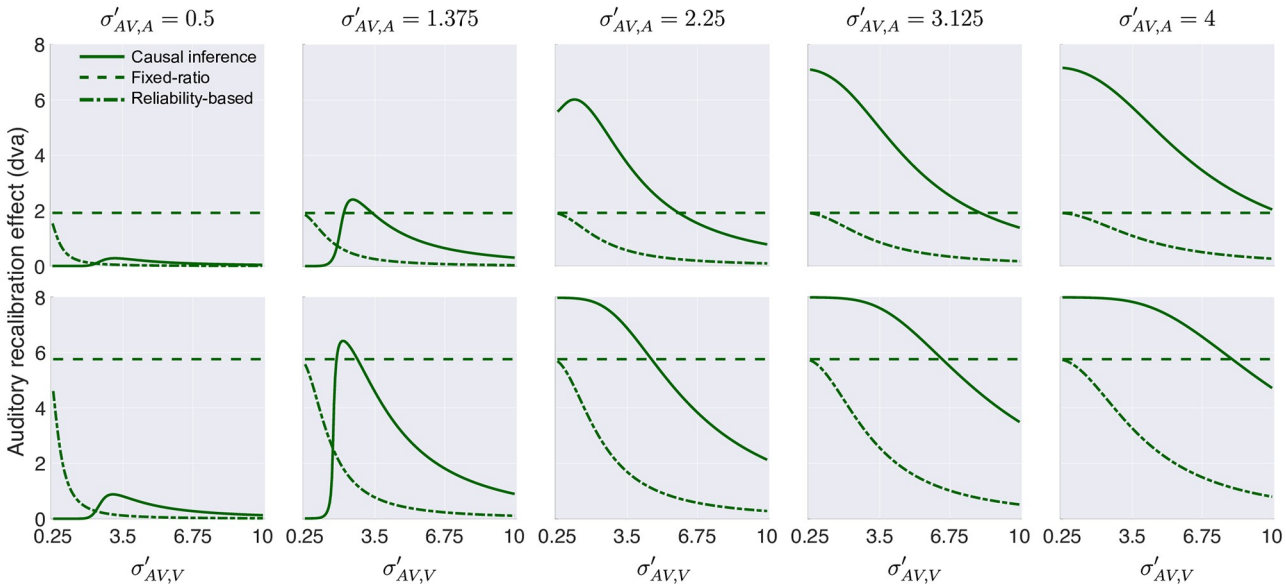

Fig 7. Simulated auditory recalibration effects based on three candidate models of cross-modal recalibration. The effect of visual (horizontal axis) and auditory (panels) stimulus reliability in bimodal trials on the amount of auditory recalibration by vision, based on causal-inference (solid line), fixed-ratio (dashed line), and reliability-based (dot-dashed line) models for two different learning rates (top row: slow; bottom row: fast).

https://doi.org/10.1371/journal.pcbi.1008877.g007

measurements. The shift for one modality is updated by a fraction of the difference between the two measurements that is proportional to the other modality's relative reliability. Across trials, the amount of recalibration depends not only on stimulus reliability, but also on a common learning rate α for both modalities. This model predicts increasing visual and decreasing auditory recalibration effects with decreasing visual stimulus reliability (Fig 7, dot-dashed line).

The fixed-ratio model. According to this model, the brain determines the amount of recalibration based on the modalities of both sensory measurements. The measurement shift for each sense is updated by a fraction of the difference between the two measurements. This fraction is determined only by modality-specific learning rates. Therefore, this model predicts modality-specific recalibration effects that are not influenced by stimulus reliability (Fig 7, dashed line).

The causal-inference model. According to this model, the brain determines the amount of recalibration for each modality based on the difference between a measurement and the corresponding location estimate. The measurement shifts are updated by a fraction of this difference, either implemented as modality-specific learning rates or as a supra-modal learning rate. In this model, cross-modal recalibration depends on stimulus reliability, modality-specific localization biases, and inference about a common cause through their influences on the location estimates. The model can predict various effects of visual reliability on auditory recalibration (Fig 7, solid line).

Modeling results

Model predictions. The causal-inference model is the only model that can capture the idiosyncratic influence of visual stimulus reliability on auditory recalibration across participants (Fig 8A). Data from 5 out of 6 participants were best fitted by the causal-inference model, either with a supra-modal learning rate (see Section S3 in S1 Appendix for model predictions) or modality-specific ones. For most participants, neither the reliability-based nor the fixed-ratio model can reproduce the diverse influence of visual stimulus reliability on auditory

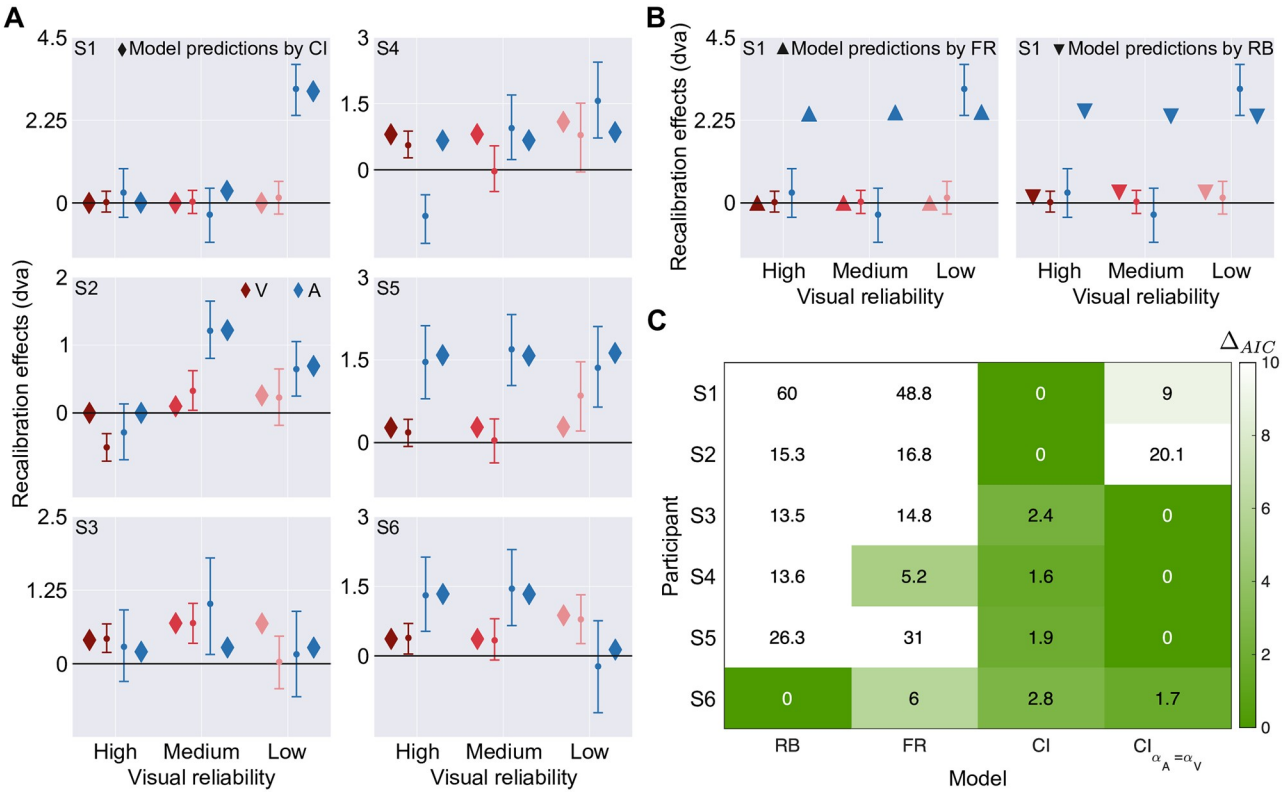

Fig 8. Model predictions and model comparison. (A) Observed (circles) and predicted (diamonds) auditory (blue) and visual (pink) final measurement shifts after the recalibration phase based on 1,000 runs using the best-fitting parameters given the causal-inference model (diamonds) as a function of visual reliability for all participants (panels). Error bars: 95% bootstrapped confidence intervals. (B) Model predictions by the fixed-ratio model (triangles in left panel) and the reliability-based model (inverted triangles in right panel) for participant S1. (C) Model comparison indices, smaller values indicate more evidence (RB = reliability-based, FR = fixed ratio, CI = causal inference, $CI_{\alpha_A=\alpha_V}$ = causal inference with supramodal learning rate).

https://doi.org/10.1371/journal.pcbi.1008877.g008

recalibration (Fig 8B; Sections S4-S5 in S1 Appendix). The three models do not differ in terms of predicting visual recalibration and modality-specific biases (Section S6 in S1 Appendix).

Model comparison. To compare model performance quantitatively, we computed the Akaike information criterion (AIC) for all three models [64] and then calculated relative model-comparison scores, Δ_{AIC}, which relate the AIC value of the best-fitting model (the model with the lowest AIC) to that of each of the other models (a high value of Δ_{AIC} indicates stronger evidence for the best-fitting model; Fig 8C). Δ_{AIC} values comparing both the reliability-based and the fixed-ratio to the causal-inference models were large for the majority of participants, revealing substantial evidence for the causal-inference model of cross-modal recalibration.

Discussion

In this study, we investigated the mechanism underlying cross-modal recalibration. To this aim, we measured the effects of visual stimulus reliability, a potential determinant of cross-modal recalibration, on audiovisual spatial recalibration. To induce recalibration, we repeatedly exposed participants to audiovisual pairs with a perceptually constant spatial discrepancy. To measure recalibration, we compared unimodal auditory and visual localization responses before and after the exposure. Auditory localization was recalibrated by vision, yet, the

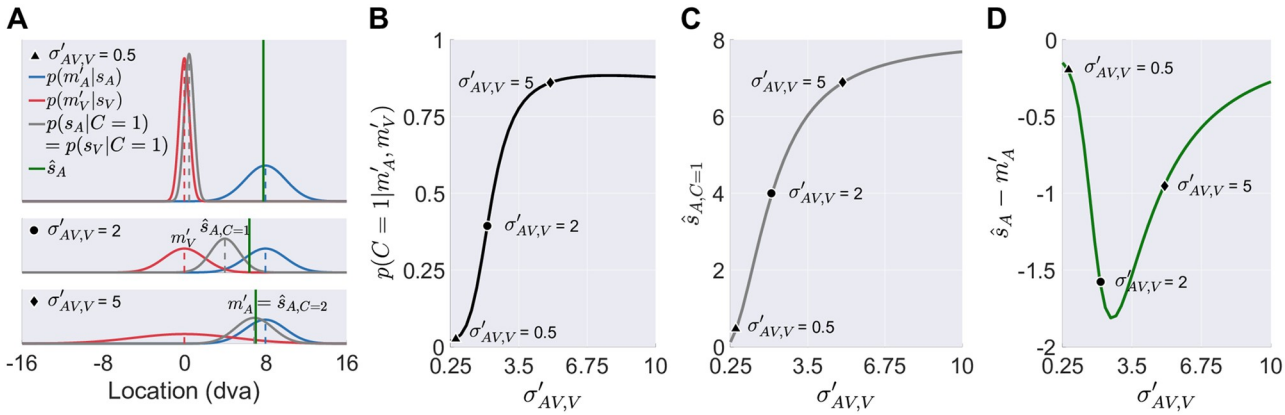

Fig 9. Cue reliability and the amount of recalibration. (A) Non-linear effect of visual stimulus reliability in bimodal trials ($\sigma'_{AV,V}$; different panels) on the auditory spatial estimate $\hat{s}_A$ (green vertical lines). Blue and red dashed vertical lines and curves: auditory and visual measurements (m'_A and m'_V in perceptual space) and likelihood functions, respectively. Blue: when separate causes are assumed, the auditory measurement and likelihood equal the auditory location estimate $\hat{s}_{A,C=2}$ and the posterior distribution over auditory locations (a flat prior over stimulus location is assumed). Grey dashed vertical lines and curves: audiovisual location estimates $\hat{s}_{A,C=1}$ and posterior distributions of audiovisual locations conditioned on a common cause. (B-D) The effect of visual reliability on the posterior probability of a common cause, $p(C = 1|m'_A, m'_V)$, the integrated location estimate, i.e., the estimate conditioned on a common audiovisual source $\hat{s}_{A,C=1}$, and the distance between auditory measurement and location estimate $\hat{s}_A - m'_A$, which directly sets the amount of recalibration.

https://doi.org/10.1371/journal.pcbi.1008877.g009

influence of visual stimulus reliability on auditory recalibration differed qualitatively across participants. To scrutinize the mechanisms of cross-modal recalibration, we compared participants' behavior to three models of recalibration: (1) a reliability-based model, which assumes that the amount of recalibration depends on the relative reliability of the cues that are in conflict, (2) a fixed-ratio model, which assumes that the amount of recalibration is fixed, dependent only on the modalities in conflict and independent of cue reliability, and (3) a causal-inference model, which ties recalibration to the percept of a cue. This percept depends on the other cue, causal inference of the two cues coming from a common source as well as cue reliabilities, and modality-specific spatial biases. Only the causal-inference model captured the idiosyncratic influences of cue reliability on cross-modal recalibration.

Only the causal-inference model captures the diverse influences of visual reliability on auditory recalibration by vision

Our results demonstrated diverse influences of visual stimulus reliability on auditory recalibration. For half of the participants, auditory recalibration was maximal at medium visual stimulus reliability. For some other participants, auditory recalibration increased with decreasing visual stimulus reliability. Neither of these patterns can be replicated by models of recalibration that assume the amount of recalibration relies directly on cue reliability [53, 61]. These models can only predict decreases in recalibration as the stimulus reliability of the other modality decreases, as has been found previously [59, 61]. Thus, the best prediction these models could produce for either monotonically or non-monotonically increasing auditory recalibration effects with decreasing visual reliability was no influence of stimulus reliability (Fig 8B, right panel). The observed influences of cue reliability on recalibration are also at odds with models of recalibration that assume the amount of recalibration relies only on the identity of the two modalities in conflict [62, 65]. These models predict no influence of stimulus reliability on recalibration. Crucially, the causal-inference model of cross-modal recalibration [32, 63] captures all the observed influences of stimulus reliability on cross-modal recalibration and is able

to replicate all previous patterns of results [61, 62] based on individual differences in cue reliability, the common-cause prior, and modality-specific spatial biases.

It is remarkable that the causal-inference model of cross-modal recalibration is capable of producing qualitatively different patterns of results for the amount of cross-modal recalibration as stimulus reliability changes [32]. Based on our experience with the model, we next provide an intuition of how individual differences in the sensory reliabilities, biases in spatial perception of both modalities, and the common-cause prior influence cross-modal recalibration according to the causal-inference model.

The role of cue reliability for cross-modal recalibration. Cue reliability influences the degree of cross-modal recalibration by influencing the posterior probability that both cues arose from a common cause as well as the integrated location estimate for the common-cause scenario. Both of them determine the final location estimate and in turn the amount of recalibration. When the visual cue is extremely reliable, the posterior probability that two discrepant cues originated from the same source is low (Fig 9B). If the common-cause scenario is unlikely, the auditory location estimate is mostly based on the estimate given separate sources and therefore located close to the auditory cue (Fig 9A, top panel), which leads to small recalibration effects.

However, the amount of auditory recalibration does not increase monotonically with decreasing visual reliability. With decreasing visual reliability, the integrated location estimate for the common-cause scenario is increasingly dominated by the auditory measurement (Fig 9C). The closer the integrated estimate is to the auditory measurement, the closer the final location estimate is to the auditory measurement, and the smaller the auditory recalibration effect will be.

Importantly, the effect of variation in reliability depends on the tested reliability range (Fig 9D). Thus, studies that use different types of stimuli will likely obtain different results regarding the influence of stimulus reliability on cross-modal recalibration. The contradictions that emerged between previous studies [61, 62] can be explained by the causal-inference model of recalibration.

The role of common cause prior assumptions for cross-modal recalibration. The common-cause prior impacts the posterior probability of a common cause and hence the amount of recalibration independent of stimulus reliability and discrepancy (Fig 10). As the common-cause prior increases, the posterior probability of a common cause increases, leading the final auditory location estimate to be farther away from the auditory measurement and hence yielding a larger measurement-shift update. We used a post-response cue in the bimodal localization task during the recalibration phase to foster audiovisual recalibration as the common-cause prior is strengthened when participants attend to both modalities [32].

Modality-specific spatial biases. Our model differs from previous causal-inference models of cross-modal perception in that we assumed the existence of modality-specific biases but not modality-specific spatial priors over stimulus location [32, 42, 66]. Both can account for the typically observed biases in localization, but modality-specific priors exert a stronger influence when visual reliability is reduced. In contrast, the influence of modality-specific biases does not vary as stimulus reliability changes. Thus, we conducted a control experiment examining whether the tendency to perceive visual stimuli as shifted towards the central fixation increased with decreasing visual reliability (Section S7 in S1 Appendix). The results showed no systematic influence of visual reliability. Thus, to avoid overfitting the model, we omitted modality-specific priors in our models and fitted only modality-specific biases.

The existence of fixed biases in the spatial perception of visual and auditory stimuli seems to be at odds with the concept of cross-modal recalibration. If discrepancies between the senses consistently lead to recalibration, why do these biases persist? Perceptual biases could reflect

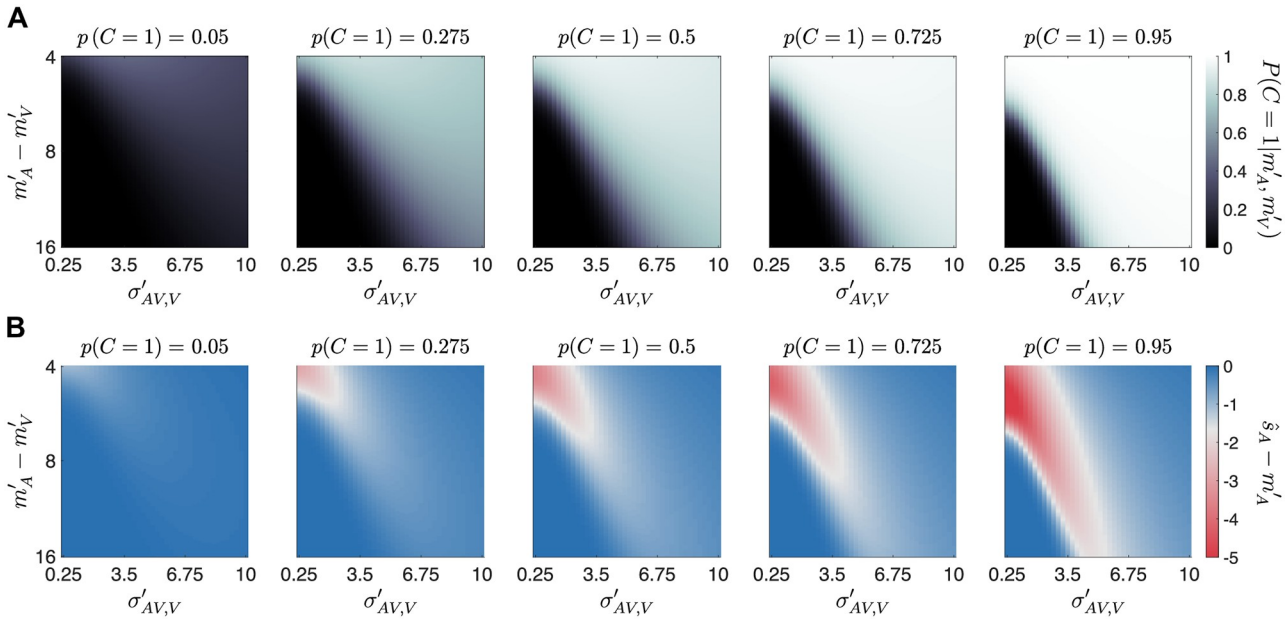

Fig 10. Determinants of recalibration. The joint effects of visual reliability in bimodal trials, $\sigma'_{AV,V}$, the distance between auditory and visual measurements in perceptual space, $m'_A - m'_V$, and the common-cause prior, $p(C = 1)$ on the posterior probability of a common cause, $p(C = 1|m'_A, m'_V)$ (panel A), and the distance between the final location estimate and the auditory measurement, $\hat{s}_A - m'_A$ (panel B), which is proportional to the amount of recalibration.

https://doi.org/10.1371/journal.pcbi.1008877.g010

adjustments for sensory discrepancies during an early sensitive period of development as has recently been shown for cross-modal biases in temporal perception [31]. After the sensitive period, it might become impossible to fully compensate for newly developing differences between the senses. In terms of our model, this would mean that a limitation to the shift-updates is set during early infancy.

Our study differs from previous spatial-recalibration studies in that we adjusted for perceptual biases in auditory relative to visual spatial perception by selecting visual locations perceptually aligned with pre-selected auditory locations. During piloting we presented stimuli with a constant physical spatial audiovisual discrepancy during the recalibration phase, and did not find significant recalibration effects. We attributed this to the modality-specific biases we had observed combined with the small spatial discrepancy used in our study. Indeed, our simulations with the causal-inference model (Fig 11, top panel) show that there are many combinations of proportional and constant biases that lead to minuscule or even negative recalibration effects when these biases are not adjusted for. Additionally, our simulations reveal a complex interaction between these biases and the spatial discrepancy. Recalibration through exposure to a constant and relatively large physical discrepancy, as done in most previous studies [43–45, 47–50, 67], would produce positive recalibration effects on average but not necessarily in all participants given that humans differ in their spatial biases. In contrast, keeping the discrepancy constant in perceptual space makes the recalibration effects less prone to individual differences in modality-specific biases and works better with smaller spatial discrepancies (Fig 11, bottom panel).

Fitting the causal-inference model

Here, we fitted, for the first time, the causal-inference model of recalibration to observed data. To achieve this, we fitted the outcome of the recalibration process rather than its build-up. The

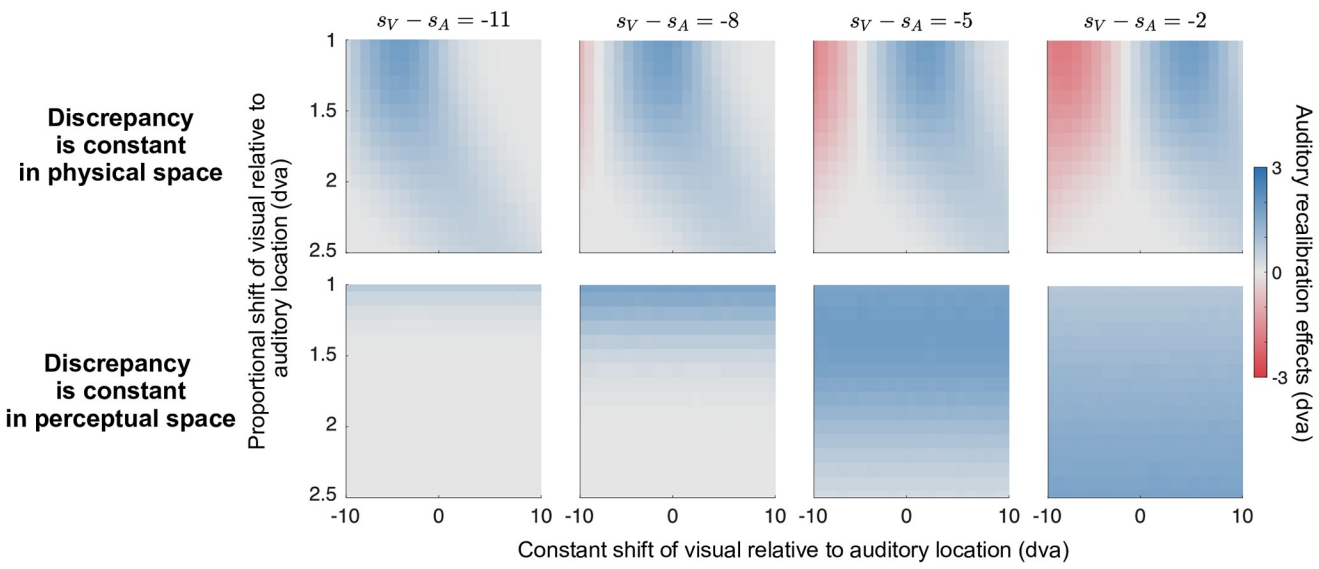

Fig 11. The influence of spatial discrepancy and modality-specific biases on the amount of auditory recalibration. Top row: The auditory recalibration effects (color key) as a function of proportional and constant shift of visual relative to auditory location when spatial discrepancy (panels) is constant in physical space (i.e., $s_V = s_A$ + spatial discrepancy). Bottom row: The auditory recalibration effects when spatial discrepancy is constant in perceptual space (i.e., visual stimulus locations are selected to adjust the perceptual biases in auditory relative to visual spatial perception, s_V = proportional shift × (s_A + spatialdiscrepancy) + constant shift).

https://doi.org/10.1371/journal.pcbi.1008877.g011

audiovisual recalibration phase itself is characterized by sequential dependence of the measurement shifts across trials and the lack of a closed-form solution for the model. Fitting such models is computationally expensive as the required number of simulations increases exponentially with each recalibration trial due to the sequential nature of the model. There might be a way to address these sequential dependencies using particle filters [68]. However, we concentrated on fitting the outcome of the recalibration process by repeatedly simulating the audiovisual recalibration phase. In this way, we obtained an approximation of the distribution of measurement-shifts given a set of parameters. The good match between the observed and predicted unimodal localization data as well as our checks of the fitting procedure (Section S8 in S1 Appendix) confirm the validity of our approach. One negative consequence of not fitting the recalibration process itself is that parameters reflecting stimulus reliability under bimodal conditions are not directly constrained by the data. We incorporated different reliabilities for unimodal and bimodal presentation conditions because previous studies indicated differences in one [32, 37, 69] or the other [70, 71] direction. Yet, we remain cautious in interpreting the estimated bimodal reliabilities in our study.

We additionally remain cautious with respect to the interpretation of other parameter estimates. We assumed a flat supra-modal prior over stimulus location as we found indications for trade-offs between modality-specific biases and a supra-modal prior over stimulus locations (Section S9.1 in S1 Appendix). As a consequence, the estimated modality-specific biases might have been underestimated and sensory reliabilities might have been overestimated. Additionally, the prior probability of a common cause and the modality-specific learning rate trade off and thus might be misestimated, because an increase in either factor can lead to a greater amount of recalibration (Section S9.2 in S1 Appendix).

Importantly, even though the possibility of biases in our parameter estimates exists, the mechanisms outlined at the beginning of the discussion explain the idiosyncratic influence of visual reliability on auditory recalibration, whereas the other models cannot qualitatively

reproduce the observed results. Thus, our conclusion that causal-inference-based percepts regulate cross-modal recalibration stands independent of the parameter estimates.

Future work might involve an experimental design that allows for better estimation of model parameters to enable a determination of the underlying cause of these individual differences and to relate them to behavior in other tasks. For example, more constraints on the experimental parameters could be obtained in a design in which unimodal trials are interspersed with recalibration trials, allowing the time course of recalibration to be measured.

Conclusion

This study examined the mechanism underlying cross-modal recalibration. To this aim, we measured audiovisual spatial recalibration while varying visual stimulus reliability. Stimulus reliability has been described as one plausible determinant the brain uses to decide which sensory modality should be recalibrated when there is a cue conflict. We found that visual stimulus reliability influenced auditory recalibration in qualitatively different ways across participants. Neither the reliability-based model nor an alternative model that assumes a fixed degree of recalibration for each modality and completely ignores stimulus reliability, could replicate the data. Yet, a causal-inference model was able to capture all the observed diverse influences of reliability on recalibration, including two patterns found in previous studies. In this model, recalibration is not based on a mere comparison of two sensory cues, but rather relates each cue to its corresponding perceptual estimate, and by doing so incorporates causal inference of a common source for a cross-modal stimulus pair as well as cue reliability and modality-specific perceptual biases into cross-modal recalibration.

Materials and methods

Ethics statement

Experimental protocols were approved by the Institutional Review Board at New York University (protocol number FY2016–595). All participants gave informed written consent prior to the beginning of the experiment and five of them were compensated $10 per hour for participation.

Participants

Six participants (three females, aged 22–29 years, mean: 25 years, six right-handed), recruited from New York University and naive to the purpose of the study, participated in the experiment. All stated that they were free of visual, auditory, or motor impairments. The data of one additional participant (female, 21 years, ambidextrous) were excluded from data analysis and model fitting due to conflicting modality-specific spatial biases found in the bimodal spatial-discrimination and unimodal localization tasks (Section S10 in S1 Appendix).

Apparatus and stimuli

The experiment was conducted in a dark and semi sound-attenuated room. Participants were seated 1 m from an acoustically transparent, white screen (1.36 × 1.02 m, 68 × 52˚ visual angle). An LCD projector (Hitachi CP-X3010N, 1024 × 768 pixels, 60 Hz) was mounted above and behind participants to project visual stimuli on the screen. The visual stimuli were clusters of 10 randomly placed low-contrast (36.55 cd/m^2) Gaussian blobs (SD: 3.6˚) added to a grey background (29.33 cd/m^2). Blob locations were drawn from a two-dimensional Gaussian distribution (vertical SD: $\sigma_y = 5.4˚$, horizontal SD: (1) $\sigma_x = 1.1˚$ for the high-reliability condition, (2) $\sigma_x = 5.4˚$ for the medium-reliability condition, and (3) $\sigma_x = 8.7˚$ for the low-reliability

condition). We recorded the centroid of the cluster rather than the center parameter of the two-dimensional Gaussian used for cluster generation as the visual stimulus location. Each visual stimulus was presented for 100 ms (6 frames), and followed by a 900 ms long backward masker (54 frames of randomly black or white checks of 4 × 4 pixels filling the screen) to erase any visual memory of the stimulus.

Behind the screen, a loudspeaker (20 W, 4Ω Full-Range Speaker, Adafruit, New York) was mounted on a sledge attached to a linear rail (1.5 m long, 23 cm above the table, 5 cm behind the screen). The rail was hung from the ceiling using elastic ropes, perpendicular to the line of sight. The position of the sledge on the rail was controlled by a microcomputer (Arduino Mega 2560; Arduino, Somerville, MA, USA). The microcomputer controlled a stepper motor that rotated a threaded rod (OpenBuilds, www.openbuildspartstore.com). This way the speaker was moved to the auditory stimulus location. The auditory stimulus was a 100 ms broadband noise burst (0–20.05 kHz, 60 dB), windowed using the first 100 ms of a sine wave with a period of 200 ms. To control audiovisual synchrony in bimodal trials, we adjusted audiovisual latencies in the presentation software and confirmed their synchrony by recording their relative latencies using a microphone and photo-diode.

We were concerned that participants might infer the position of the speaker from the sounds produced by sledge movements. We tried to foil this strategy by playing a masking sound from an additional speaker behind the center of the screen during each movement of the speaker. The masking sound (55 dB) was a recording of a randomly chosen speaker movement plus white noise. Additionally, the speaker moved from its last position to the target location through a stopover location. The stopover was randomly chosen under the constraint that the total distances the speaker moved were approximately equal across trials. We carried out a control experiment, which indicated that participants could not infer the speaker position based on sounds arising from the speaker movements (Section S11 in S1 Appendix).

Responses were given using a numeric keypad in spatial-discrimination tasks, and a pointing device in localization tasks. The pointing device was custom built using a potentiometer (Uxcell 10KΩ Linear Taper Rotary Potentiometer) with a plastic ruler (5 × 17 cm) securely fixed perpendicular to the shaft of the potentiometer. Participants placed their hands on either side of the ruler and rotated it so that it pointed at the perceived location of the stimulus. A visual cursor (an 8 × 8 pixel white square) was displayed to indicate the selected location. The pointing device was covered by a black box (42 × 30 × 15 cm) so that participants could not see their hands while using the device. A foot pedal was placed on the floor and used to confirm the current position of the pointing device as the response.

Stimulus presentation, speaker movement, and response collection were controlled by a laptop PC running MATLAB R2017b (MathWorks, Natick, MA, USA). Visual stimuli were presented using the Psychophysics Toolbox [72–74].

Procedure

In the beginning of the study, participants completed a unimodal spatial-discrimination task, once for the auditory stimulus, and once for each of the three visual stimuli (low, medium, and high spatial uncertainty). Participants then completed a bimodal spatial-discrimination task followed by a pointing practice task. The last part of the study consisted of six recalibration sessions, one for each condition (2 recalibration directions × 3 reliability levels). Each session started with a unimodal localization task, followed by an audiovisual recalibration phase in which participants localized one modality of a spatially discrepant audiovisual stimulus, and ended with a repetition of the unimodal localization task.

Unimodal spatial-discrimination task. Participants' spatial-discrimination thresholds were measured using a 2-interval, forced-choice (2IFC) procedure. In each trial, a standard and a test stimulus were presented in random order. Each stimulus was preceded by a fixation cross, presented at the center of the screen for 1,000 ms, followed by a 2,000 ms-long period of blank screen in which the loudspeaker moved to its position. The actual stimulus lasted 100 ms, followed by either a 900 ms-long backward masker (visual stimuli) or a blank screen for 900 ms (auditory stimuli). After the second stimulus period was over, a response probe was displayed and participants indicated by button press which interval contained the stimulus that was farther to the right (Fig 1B). Visual feedback was provided for 500 ms immediately after the response was given. The inter-trial interval was 500 ms.

The standard stimulus was located straight-ahead at the center of the screen; the location of the test stimulus was controlled by four interleaved staircases, two of which had the test stimulus start at 12.5˚ (to the right of straight-ahead), and the other two at −12.5˚ (12.5˚ to the left of straight-ahead). For the two staircases starting at one side, one followed the two-down-one-up rule (with down being defined as moving the test stimulus leftwards) converging to a probability of 71% [75] of perceiving the test stimulus as farther to the right than the standard stimulus; the other staircase followed the one-down-two-up rule converging to a probability of 29%. The initial step size was 1.9˚, decreased to 1.0˚ after the first staircase reversal and to 0.5˚ after the third reversal. Each staircase consisted of 40 trials, resulting in a total of 160 trials. To improve the estimation of the lapse rate, a test stimulus was presented distant from the center (±12.5˚) once every 10 trials. A total of 176 trials was evenly split into 4 blocks.

Participants completed the spatial-discrimination task for each of the three levels of visual stimulus reliability in random order, the auditory stimulus was always tested last. Participants took about an hour to complete one stimulus condition; typically, they spread all four stimulus conditions across two days.

Bimodal spatial-discrimination task. Participants' biases in auditory relative to visual spatial perception were measured using a spatial 2IFC procedure. An auditory and a high-reliability visual stimulus were presented in random order and participants indicated by button press whether the visual stimulus was located to the left or right of the auditory stimulus. The procedure was otherwise identical to that of the unimodal spatial-discrimination task (Fig 3A). No feedback was provided.

The auditory stimulus was presented at one of four locations (±2.5 or ±7.5˚). We used the adaptive staircase procedure to effectively sample the visual stimulus space for each participant as the range of meaningful stimulus locations varied considerably across participants due to individual perceptual biases. Specifically, the location of the visual stimulus was controlled by eight interleaved staircases, two for each of the four auditory locations. Of the two staircases per auditory location, one started the test stimulus at 15˚ relative to the auditory location, and the other one at −15˚. Staircases with the visual stimulus starting from the right of the auditory stimulus followed the one-down-two-up rule, converging to a probability of 29% of choosing the visual as to the right of the auditory stimulus. Staircases with the visual stimulus starting from the left of the auditory stimulus followed the two-down-one-up rule, converging to a probability of 71%. Staircase step size was updated as described above. Each staircase consisted of 36 trials. Trials with the visual stimulus being located at ±15˚ relative to the auditory stimulus were inserted once every nine trials, resulting in a total of 320 trials. The session was divided into six blocks. Usually, participants took about two hours to complete all trials.

Pointing practice task. Participants' localization precision independent of spatial perception was measured using a localization task with visual stimuli of maximal spatial reliability. In each trial, a white square (8 × 8 pixels ≈ 0.6˚ × 0.6˚) was displayed on the screen for 100 ms. The stimulus was followed by a 900 ms-long backward masker and 500 ms of blank screen.

Subsequently, the response cursor, a green square of the same size as the white square, appeared on the screen. Participants used the pointing device to move the cursor to the stimulus position, and confirmed their response with the footpedal. Response times were unrestricted. The cursor location was shown during adjustment, but error feedback was not provided. There were eight possible horizontal positions for the stimulus, evenly spaced from -17.5 to 17.5˚ in steps of 5˚. Each stimulus location was visited 30 times in random order, resulting in a total of 240 trials. The inter-trial interval was 500 ms. This experiment took 30 minutes to complete and was typically administered after the bimodal spatial-discrimination task on the same day.

Unimodal localization task (pre- and post-recalibration phase). Participants' baseline and post-recalibration spatial perception were measured using a unimodal spatial-localization task. In each trial, either an auditory or a visual stimulus was presented, again preceded by a fixation cross and a blank screen. The inter-trial interval was 100 ms, otherwise timing was identical to that of the discrimination tasks. Auditory stimulus locations were the same as in the bimodal spatial-discrimination task; visual stimulus locations were the four locations that were identified as perceptually co-located with those four auditory locations using the bimodal spatial-discrimination task. Participants again responded by moving a visual cursor to the stimulus location (Fig 5A). There was no time limit for the response. The location of the visual cursor was shown during the adjustment, but feedback about the localization error was not provided. Each of the four target locations per modality was tested 12 times, resulting in a total of 96 trials administered in pseudorandom order. These trials were split into four blocks. Usually participants took about 25 min to complete all 96 trials; they did so once at the beginning of the session (pre-recalibration phase) and once again after the recalibration phase (post-recalibration phase).

Bimodal localization task (recalibration phase). During the audiovisual recalibration phase, participants were presented with temporally synchronous but spatially discrepant audiovisual stimuli. We asked them to localize either the auditory or the visual component, with the localization modality cued after stimulus presentation (Fig 5B), a procedure that has been associated with larger recalibration effects than non-spatial tasks [32]. All other trial parameters were identical to the unimodal localization task. Three audiovisual stimulus pairs were chosen such that an auditory stimulus location was paired with the visual location perceived as aligned with the auditory stimulus location to its left in the visual-left-of-auditory condition and the auditory stimulus location to its right in the visual-right-of-auditory condition (Fig 5C). Each of the three audiovisual pairs was repeated 40 times in random order, resulting in a total of 120 trials. These trials were split into four blocks. Usually, participants took about 30 min to complete all 120 trials. A full session (pre-recalibration, recalibration, and post-recalibration phase) took 80 min. Participants completed the six sessions on separate days.

Data preparation and statistical analysis

Unimodal spatial-discrimination task. For the unimodal spatial-discrimination task, the data were coded as the probability of identifying the test stimulus as located farther to the right than the standard stimulus as a function of test stimulus location separately for each stimulus type (Fig 2A). These data were fitted with a cumulative Gaussian distribution centered at 0 and with a lapse rate constrained to be less than or equal to 6% [76]. The JND was calculated as half the distance between the stimulus locations corresponding to probabilities of 0.25 and 0.75 according to the fitted cumulative Gaussian distribution (unscaled by the lapse rate). To measure goodness of fit, we computed adjusted R^2 values based on the binned data (bin

size = 1.8˚). To derive error bars, we randomly resampled the raw data with replacement 1,000 times, fitted psychometric functions to each resampled dataset, calculated the JND, and took the 2.5 and 97.5 percentiles of the 1,000 JNDs as the bootstrapped confidence interval.

Bimodal spatial-discrimination task. Data from the bimodal spatial-discrimination task were coded as the probability of identifying the visual stimulus as located to the right of the auditory stimulus as a function of visual stimulus location. We fitted four cumulative Gaussian distributions to these data, one for each auditory standard stimulus. Again, we included a lapse rate, constrained to be less than or equal to 6% [76]. Adjusted R^2 values were calculated based on binned data (bin size = 3˚). The point of subjective equality (PSE) was defined as the visual stimulus location corresponding to a probability of 0.5 according to the unscaled psychometric function. The four PSEs were modeled as a linear function of auditory stimulus location. From this linear regression of the PSEs, we computed the locations of the visual stimulus perceived as co-located with the four auditory locations. In the subsequent unimodal and bimodal spatial-localization tasks, we presented visual stimuli at these locations rather than those directly indicated by the PSEs to reduce effects of random noise during the bimodal spatial-discrimination task. 95% confidence intervals for each parameter were obtained as before.

Pointing practice task. Data from the pointing practice task, used to measure localization response precision, were not statistically analyzed but were filtered before the model fitting. For each stimulus location, we *z*-transformed the data by subtracting the mean localization response per stimulus location and then dividing by the standard deviation of all demeaned responses. Localization responses with a *z*-score outside of [−3, 3] were identified as outliers (0–1.67% of trials) and excluded from the model fitting.

Unimodal localization task (pre- and post-recalibration phase). Data from the unimodal localization task were filtered separately for each modality, stimulus location and phase. To compute the means for the z-transformation, auditory and visual localization responses in the pre-recalibration phase were pooled across all six sessions, as the day of testing should not influence localization performance. In contrast, the means of auditory and visual localization responses in the post-recalibration phase were calculated separately for each of the six conditions as each recalibration condition should influence localization differently. Then, we computed the standard deviation of the demeaned auditory localization responses pooled across all six session and both phases. The standard deviation of demeaned visual localization responses was calculated separately for each visual-reliability condition given that stimulus reliability should influence localization precision. Localization responses identified as outliers (*z*-scores outside of [−3, 3]; 0.78–1.82% of trials) were excluded from all further analyses.

For statistical analysis and display, we summarized localization responses as a linear function of stimulus location, separately for each modality and phase. For each modality, localization responses in the pre-recalibration phase were pooled across all six sessions (visual reliability should influence the precision but not the accuracy of the localization responses; Section S7 in S1 Appendix). Responses from the post-recalibration phase were regressed separately for each condition, because each condition should influence localization differently. The seven regression lines per modality were fit with the constraint that all have the same slope. The amount of recalibration of one modality by the other was defined as the distance between the intercepts of the regression lines for pre- and post-recalibration localization responses. It was coded as positive if localization responses in the post-recalibration phase were shifted to compensate for the audiovisual discrepancy in the preceding recalibration phase. For the statistical analysis, we calculated the amount of recalibration for each modality and recalibration condition (2 recalibration directions × 3 reliability levels). To derive confidence intervals, localization responses were resampled, separately for each location, task, session, and modality.

Data from the bimodal spatial-localization task conducted during the audiovisual recalibration phase were not analyzed. Data analysis was done using Python 3.7, R 4.0.2, and MATLAB 2019a.

Models of audiovisual recalibration

In this section we lay out the definition of recalibration underlying all models reported here and then describe the recalibration process during the audiovisual recalibration phase according to each of the models. Finally, we provide a formalization of each of the tasks used to constrain the model parameters followed by the details of how the models were fit to the data.

Definition of recalibration

Each stimulus at location s in the world leads to a sensory measurement m' in an observer's brain. This measurement is corrupted by Gaussian-distributed sensory noise. Thus, with repeated presentations of stimuli at location s, the sensory measurements correspond to scattered spatial locations $m' \sim \mathcal{N}(s', \sigma'^2)$. The variability of the measurements is determined by the stimulus reliability $1/\sigma'^2$. To allow integration of information from different modalities, measurements are remapped into a common internal reference frame. Hence, the measurement distribution is centered on s', the remapped location of s. As part of the remapping process, spatial discrepancies between the senses are accounted for by shifting the measurements by a modality-specific amount Δ. We model recalibration as the process of updating these shifts following each encounter with a cross-modal stimulus pair [32, 63] and probed this updating process by misaligning the physical visual and auditory stimuli to create an artificial sensory discrepancy.

We assumed that observers were calibrated as far as possible at the beginning of each experimental session. Thus, the remapped stimulus locations in internal space were understood as linear functions of the physical stimulus locations s_A and s_V, that is, $s'_A = a_A s_A + b_A$ and $s'_V = a_V s_V + b_V$. However, given that we can only measure relative biases there is no way to empirically isolate the remapping of one modality. As a consequence and without loss of generality, we set s'_V to be equal to s_V (i.e., $a_V = 1$ and $b_V = 0$).

We use the variables Δ_{A_i} and Δ_{V_i} to exclusively capture the update in measurement shifts after encounters with the spatially discrepant audiovisual stimulus pairs with visual reliability i ($i \in \{1, 2, 3\}$) during the audiovisual recalibration phase. In addition, we assumed that Δ_{A_i} and Δ_{V_i} are updated in every trial of the phase, and thus the location-independent shifts at the end of the task can be written as the sum of the initial shifts and the shift updates over 120 trials, that is $s'_A = a_A s_A + b_A + \Delta_{A_i}(121)$ and $s'_V = s_V + \Delta_{V_i}(121)$. The final shifts accumulated after 120 recalibration trials were assumed be maintained throughout the subsequent post-recalibration task as observers were not exposed to spatially aligned audiovisual pairs after the recalibration phase (Section S12 in S1 Appendix).

We further assumed that stimulus reliability differed between unimodal ($1/{\sigma'_A}^2$) and bimodal ($1/{\sigma'_{AV,A}}^2$) stimulus presentations. Note that we denote the visual-reliability condition for variables associated with the auditory modality (i.e., with a subscript A_i) when the value of that variable can be impacted by visual measurements (e.g., shifts Δ_{A_i} or, below, sensory estimates $\hat{s}'_{A_i}$), but not otherwise (e.g., measurements m'_A or measurement variances ${\sigma'_A}^2$ and ${\sigma'_{AV,A}}^2$).

Models of the recalibration process

The reliability-based model of cross-modal recalibration. According to this model, each modality should be recalibrated in the direction of the other modality by an amount that is proportional to the other modality's relative reliability [53]. In other words, after every trial, the measurement shifts, Δ_{A_i} and Δ_{V_i}, are updated in the direction of the discrepancy between the visual and auditory measurements by an amount proportional to the two modalities' relative reliabilities as follows:

$$\Delta_{A_i}(t+1) = \Delta_{A_i}(t) + \alpha w_i (m'_{V_i,l_V(t)} - m'_{A,l_A(t)}), \tag{1}$$

where

$$w_i = \frac{{\sigma'_{AV,V_i}}^{-2}}{{\sigma'_{AV,V_i}}^{-2} + {\sigma'_{AV,A}}^{-2}} \tag{2}$$

and analogously

$$\Delta_{V_i}(t+1) = \Delta_{V_i}(t) + \alpha(1 - w_i)(m'_{A,l_A(t)} - m'_{V_i,l_V(t)}), \tag{3}$$

where l_A and l_V index the auditory and visual locations (l_A, $l_V \in \{1, 2, 3, 4\}$), α denotes a supra-modal learning rate, and t denotes trial number.

The fixed-ratio model of cross-modal recalibration. According to this model, after every trial, Δ_{A_i} and Δ_{V_i} are updated in the direction of the discrepancy between the visual and auditory measurements by a fixed ratio of this discrepancy. The ratio of the update depends solely on the identity of the modality and thus is independent of stimulus reliability [62]. Δ_{A_i} and Δ_{V_i} are updated according to the following equations:

$$\Delta_{A_i}(t+1) = \Delta_{A_i}(t) + \alpha_A (m'_{V_i,l_V(t)} - m'_{A,l_A(t)}) \tag{4}$$

and

$$\Delta_{V_i}(t+1) = \Delta_{V_i}(t) + \alpha_V (m'_{A,l_A(t)} - m'_{V_i,l_V(t)}), \tag{5}$$

where α_A and α_V are modality-specific learning rates.

The causal-inference model of recalibration. In this model, the shift updates are determined by the discrepancy between a measurement and the corresponding perceptual estimate, $\hat{s}'_{A_i,l_A(t)} - m'_{A,l_A(t)}$ and $\hat{s}'_{V_i,l_V(t)} - m'_{V_i,l_V(t)}$ [63] for each modality. The spatial discrepancy between auditory and visual measurements and the relative reliabilities of both stimuli have indirect influence on the shift updates by means of their influence on the location estimates, $\hat{s}'_{A_i,l_A(t)}$ and $\hat{s}'_{V_i,l_V(t)}$. Additionally, the location estimates and thus the shift updates are contingent on the degree to which the brain infers a common cause or separate causes for the two measurements [30, 63].

The location estimates are a mixture of two conditional location estimates, one for each causal scenario (common audiovisual source, $C = 1$, or different auditory and visual sources, $C = 2$). In the case of a common source, the location estimate of the audiovisual stimulus pair, $\hat{s}'_{A_i,l_A(t),C=1} = \hat{s}'_{V_i,l_V(t),C=1}$, equals the reliability-weighted average of the measurements $m'_{A,l_A(t)}$, $m'_{V_i,l_V(t)}$ and the mean μ'_P of an internal, Gaussian-shaped, supra-modal prior across stimulus

locations with variance σ'^{2}_{P}:

$$\hat{s}'_{A_i,l_A(t),C=1} = \hat{s}'_{V_i,l_V(t),C=1} = \frac{m'_{A,l_A(t)}\sigma'^{-2}_{AV,A} + m'_{V_i,l_V(t)}\sigma'^{-2}_{AV,V_i} + \mu'_P\sigma'^{-2}_{P}}{\sigma'^{-2}_{AV,A} + \sigma'^{-2}_{AV,V_i} + \sigma'^{-2}_{P}}. \quad (6)$$

In the case of two separate sources, the location estimates of the auditory and the visual stimulus, $\hat{s}'_{A_i,l_A(t),C=2}$ and $\hat{s}'_{V_i,l_V(t),C=2}$, are equal to the reliability-weighted averages of $m'_{A,l_A(t)}$ and μ'_P for the auditory estimate, and $m'_{V_i,l_V(t)}$ and μ'_P for the visual estimate, respectively:

$$\hat{s}'_{A_i,l_A(t),C=2} = \frac{m'_{A,l_A(t)}\sigma'^{-2}_{AV,A} + \mu'_P\sigma'^{-2}_{P}}{\sigma'^{-2}_{AV,A} + \sigma'^{-2}_{P}} \quad (7)$$

and

$$\hat{s}'_{V_i,l_V(t),C=2} = \frac{m'_{V_i,l_V(t)}\sigma'^{-2}_{AV,V_i} + \mu'_P\sigma'^{-2}_{P}}{\sigma'^{-2}_{AV,V_i} + \sigma'^{-2}_{P}}. \quad (8)$$

The final location estimates are derived by model averaging (see alternative decision strategy in Section S3 in S1 Appendix). Specifically, the final location estimate $\hat{s}'_{A_i,l_A(t)}$ is the average of the conditional location estimates, $\hat{s}'_{A_i,l_A(t),C=1}$ and $\hat{s}'_{A_i,l_A(t),C=2}$, with each estimate weighted by the posterior probability of its causal structure:

$$\begin{aligned}\hat{s}'_{A_i,l_A(t)} = \quad & \hat{s}'_{A_i,l_A(t),C=1}P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)}) \\ & +\hat{s}'_{A_i,l_A(t),C=2}(1-P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)}))\end{aligned} \quad (9)$$

and analogously for the visual location estimate:

$$\begin{aligned}\hat{s}'_{V_i,l_V(t)} = \quad & \hat{s}'_{V_i,l_V(t),C=1}P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)}) \\ & +\hat{s}'_{V_i,l_V(t),C=2}(1-P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)})).\end{aligned} \quad (10)$$

The posterior probability of a common source for the auditory and visual measurements in trial t, $P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)})$, is proportional to the product of the likelihood of a common source for these measurements in trial t and the prior probability of a common source for visual and auditory measurements in general, $P(C=1)$:

$$P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)}) = \frac{P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=1)P(C=1)}{P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=1)P(C=1) + P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=2)(1-P(C=1))}. \quad (11)$$

The posterior probability of two separate sources, one for the auditory and one for the visual measurement, is $1-P(C=1|m'_{A,l_A(t)}, m'_{V_i,l_V(t)})$.

The likelihood of a common source of the visual and auditory measurements in trial t, $P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=1)$, is the product of the likelihood of the internally represented audiovisual stimulus location s'_{AV} given the auditory measurement, $m'_{A,l_A(t)}$, and the visual measurement $m'_{V_i,l_V(t)}$, and the supra-modal prior, integrated over all possible remapped audiovisual

stimulus locations s'_{AV} [30]:

$$P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=1) = \int P(m'_{A,l_A(t)}|s'_{AV})P(m'_{V_i,l_V(t)}|s'_{AV})P(s'_{AV})ds'_{AV}$$
$$= \frac{1}{2\pi\sqrt{\sigma'^2_{AV,A}\sigma'^2_{AV,V_i} + \sigma'^2_{AV,A}\sigma'^2_P + \sigma'^2_{AV,V_i}\sigma'^2_P}} \times \quad (12)$$
$$\exp\left[-\frac{1}{2}\frac{(m'_{A,l_A(t)}-m'_{V_i,l_V(t)})^2\sigma'^2_P + (m'_{A,l_A(t)}-\mu'_P)^2\sigma'^2_{AV,V_i} + (m'_{V_i,l_V(t)}-\mu'_P)^2\sigma'^2_{AV,A}}{\sigma'^2_{AV,A}\sigma'^2_{AV,V_i} + \sigma'^2_{AV,A}\sigma'^2_P + \sigma'^2_{AV,V_i}\sigma'^2_P}\right].$$

The likelihood of different sources for the visual and auditory measurements in trial t, $P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=2)$ is the product of the likelihood of internally represented auditory and visual stimulus locations s'_A and s'_V given the auditory measurement, $m'_{A,l_A(t)}$, and the visual measurement $m'_{V_i,l_V(t)}$, and the supra-modal prior. Given that the measurements in this causal scenario stem from different sources, the product is integrated over all possible, remapped visual and auditory stimulus locations, s'_V and s'_A:

$$P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|C=2) = \int\int P(m'_{A,l_A(t)}, m'_{V_i,l_V(t)}|s'_A, s'_V)P(s'_A, s'_V)ds'_A ds'_V$$
$$= \left(\int P(m'_{A,l_A(t)}|s'_A)P(s'_A)ds'_A\right)\left(\int P(m'_{V_i,l_V(t)}|s'_V)P(s'_V)ds'_V\right)$$
$$= \frac{1}{2\pi\sqrt{(\sigma'^2_{AV,A} + \sigma'^2_P)(\sigma'^2_{AV,V_i} + \sigma'^2_P)}} \times \quad (13)$$
$$\exp\left[-\frac{1}{2}\left(\frac{(m'_{A,l_A(t)} - \mu'_P)^2}{\sigma'^2_{AV,A} + \sigma'^2_P} + \frac{(m'_{V_i,l_V(t)} - \mu'_P)^2}{\sigma'^2_{AV,V_i} + \sigma'^2_P}\right)\right].$$

The updates of the shifts are scaled by two modality-specific learning rates, α_A and α_V [63]:

$$\Delta_{A_i}(t+1) = \Delta_{A_i}(t) + \alpha_A(\hat{s}'_{A_i,l_A(t)} - m'_{A,l_A(t)}) \quad (14)$$

and

$$\Delta_{V_i}(t+1) = \Delta_{V_i}(t) + \alpha_V(\hat{s}'_{V_i,l_V(t)} - m'_{V_i,l_V(t)}). \quad (15)$$

We also tested a version of the model with one supra-modal learning rate ($\alpha = \alpha_A = \alpha_V$).

Formalization of the tasks

Unimodal spatial-discrimination task. The unimodal spatial-discrimination task was conducted to estimate the one auditory $1/\sigma'_A 2$ and three visual $1/\sigma'_{V_i} 2$ stimulus reliabilities under unimodal presentation conditions as well as to constrain the estimates of the variable bias a_A introduced by the remapping process. We begin by describing the auditory version. The standard stimulus was presented straight ahead, at location $s_{A,0}$, and the test stimulus was presented at one of N_A locations, $s_{A,n}$, determined by an adaptive procedure. For each pair, the probability, $p_{A,n}$, of estimating the test stimulus to be located to the right of the standard stimulus is a function of the physical distance between the two stimuli.

We assume that the observer makes the decision by comparing the internal location estimates $\hat{s}'_{A,n}$ and $\hat{s}'_{A,0}$,

$$p_{A,n} = P(\hat{s}'_{A,n} > \hat{s}'_{A,0}) = P(\hat{s}'_{A,n} - \hat{s}'_{A,0} > 0). \tag{16}$$

To further specify $p_{A,n}$, we have to derive the probability distributions of the internal location estimates. Each physical stimulus at location $s_{A,n}$ results in an internal measurement, $m'_{A,n}$. The measurement distribution is Gaussian ($m'_{A,n} \sim \mathcal{N}(s'_{A,n}, {\sigma'_A}2)$) and for a given measurement the estimate of the remapped location of the stimulus is the average of the measurement and the mean of the spatial prior, μ'_P, each weighted by their relative reliabilities ($\hat{s}'_{A,n} = \frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2}+{\sigma'_P}^{-2}} m'_{A,n} + \frac{{\sigma'_P}^{-2}}{{\sigma'_A}^{-2}+{\sigma'_P}^{-2}} \mu'_P$). Thus, the probability distribution of the location estimates of a test stimulus is

$$\hat{s}'_{A,n} \sim \mathcal{N}(\mu_{\hat{s}'_{A,n}}, \sigma^2_{\hat{s}'_A}) \tag{17}$$

where

$$\mu_{\hat{s}'_{A,n}} = \frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2} + {\sigma'_P}^{-2}} s'_{A,n} + \frac{{\sigma'_P}^{-2}}{{\sigma'_A}^{-2} + {\sigma'_P}^{-2}} \mu'_P \text{ and } \sigma^2_{\hat{s}'_A} = \left(\frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2} + {\sigma'_P}^{-2}}\right)^2 {\sigma'_A}^2 \tag{18}$$

The probability distribution of the difference between the two location estimates $\hat{s}'_{A,n}$ and $\hat{s}'_{A,0}$ is

$$\hat{s}'_{A,n} - \hat{s}'_{A,0} \sim \mathcal{N}(\mu_{\hat{s}'_{A,n}} - \mu_{\hat{s}'_{A,0}}, 2\sigma^2_{\hat{s}'_A}), \tag{19}$$

where $\mu_{\hat{s}'_{A,0}} = \frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2}+{\sigma'_P}^{-2}} s'_{A,0} + \frac{{\sigma'_P}^{-2}}{{\sigma'_A}^{-2}+{\sigma'_P}^{-2}} \mu'_P$. Taken together, the probability of perceiving an auditory test stimulus at location $s_{A,n}$ to the right of an auditory standard stimulus at location $s_{A,0}$ is

$$\begin{aligned} p_{A,n} &= P(\hat{s}'_{A,n} - \hat{s}'_{A,0} > 0) \\ &= 1 - \Phi(0; \mu_{\hat{s}'_{A,n}} - \mu_{\hat{s}'_{A,0}}, 2\sigma^2_{\hat{s}'_A}) \\ &= 1 - \Phi\left(0; \frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2} + {\sigma'_P}^{-2}} (s'_{A,n} - s'_{A,0}), 2\left(\frac{{\sigma'_A}^{-2}}{{\sigma'_A}^{-2} + {\sigma'_P}^{-2}}\right)^2 {\sigma'_A}^2\right) \\ &= \Phi(s'_{A,n} - s'_{A,0}; 0, 2{\sigma'_A}^2), \end{aligned} \tag{20}$$

where $\Phi(x;\mu, \sigma^2)$ is the cumulative Gaussian distribution.

However, as experimenters we only have access to response probabilities as a function of the stimulus locations in physical space. Given that the remapped location of s'_A is a function of the physical stimulus location s_A, we can rewrite Eq 20 as

$$\begin{aligned} p_{A,n} &= \Phi((a_A s_{A,n} + b_A) - (a_A s_{A,0} + b_A); 0, 2{\sigma'_A}^2) \\ &= \Phi(s_{A,n} - s_{A,0}; 0, 2a_A^{-2}{\sigma'_A}^2), \end{aligned} \tag{21}$$

and analogously,

$$p_{V_i,n} = \Phi(s_{V_i,n} - s_{V_i,0}; 0, 2{\sigma'_{V_i}}^2). \tag{22}$$

Finally, the model includes occasional response lapses (i.e., random button presses) at rate λ, so that the probability of reporting the test stimulus as located farther to the right than the

standard ($r_{A,n} = 1$) is

$$p_{r_{A,n}=1} = 0.5\lambda_A + (1-\lambda_A)p_{A,n} \text{ and } p_{r_{V_i,n}=1} = 0.5\lambda_{V_i} + (1-\lambda_{V_i})p_{V_i,n}. \quad (23)$$

Bimodal spatial-discrimination task. The bimodal spatial-discrimination task was conducted to estimate the relative bias of auditory compared to visual spatial perception, i.e., to estimate a_A and b_A. Auditory stimuli were presented at four different locations in physical space $s_{A,l}$, where l indexes the auditory location. Guided by a staircase procedure, on each trial t, an auditory stimulus at location $s_{A,l}$ was paired with a visual stimulus of high spatial reliability ($i = 1$) at one of N test locations $s_{V,n}$, where n indexes the finer grid of locations of visual stimuli that were presented during the task. For each pair, the model predicts $p_{l,n}$, the probability of judging the visual stimulus at location $s_{V,n}$ as to the right of the auditory stimulus at location $s_{A,l}$.

Analogous to the unimodal spatial-discrimination task, we specify the probability distributions of the internal auditory and visual location estimates as

$$\hat{s}'_{A,l} \sim \mathcal{N}(\mu_{\hat{s}'_{A,l}}, \sigma^2_{\hat{s}'_A}) \text{ and } \hat{s}'_{V_1,n} \sim \mathcal{N}(\mu_{\hat{s}_{V_1,n}}, \sigma^2_{\hat{s}_{V_1}}), \quad (24)$$

where

$$\mu_{\hat{s}'_{A,l}} = \frac{\sigma'_A{}^{-2} s'_{A,l} + \sigma'_P{}^{-2}\mu'_P}{\sigma'_A{}^{-2} + \sigma'_P{}^{-2}} = \frac{\sigma'_A{}^{-2}(a_A s_{A,l} + b_A) + \sigma'_P{}^{-2}\mu'_P}{\sigma'_A{}^{-2} + \sigma'^{-2}_P}, \quad (25)$$

$$\mu_{\hat{s}'_{V_1,n}} = \frac{\sigma'_{V_1}{}^{-2} s'_{V,n} + \sigma'_P{}^{-2}\mu'_P}{\sigma'_{V_1}{}^{-2} + \sigma'_P{}^{-2}} = \frac{\sigma'_{V_1}{}^{-2} s_{V,n} + \sigma'_P{}^{-2}\mu'_P}{\sigma'_{V_1}{}^{-2} + \sigma'_P{}^{-2}}, \quad (26)$$

$$\sigma^2_{\hat{s}'_A} = \frac{\sigma'_A{}^{-2}}{(\sigma'_A{}^{-2} + \sigma'_P{}^{-2})^2} \text{ and } \sigma^2_{\hat{s}'_{V_1}} = \frac{\sigma'_{V_1}{}^{-2}}{(\sigma'_{V_1}{}^{-2} + \sigma'_P{}^{-2})^2}. \quad (27)$$

The probability of the observer perceiving a visual stimulus at physical location $s_{V,n}$ as located to the right of an auditory stimulus at physical location $s_{A,l}$ is thus:

$$p_{l,n} = P(\hat{s}'_{V_1,n} > \hat{s}'_{A,l}) = P(\hat{s}'_{V_1,n} - \hat{s}'_{A,l} > 0). \quad (28)$$

The distribution of this difference is:

$$\hat{s}'_{V_1,n} - \hat{s}'_{A,l} \sim \mathcal{N}(\mu_{\hat{s}'_{V_1,n}} - \mu_{\hat{s}'_{A,l}}, \sigma_{\hat{s}'_{V_1}}{}^2 + \sigma_{\hat{s}'_A}{}^2). \quad (29)$$

The probability of perceiving the visual stimulus to the right of the auditory can then be expressed as

$$p_{l,n} = 1 - \Phi(0; \mu_{\hat{s}'_{V_1,n}} - \mu_{\hat{s}'_{A,l}}, \sigma^2_{\hat{s}'_{V_1}} + \sigma^2_{\hat{s}'_A}). \quad (30)$$

As in the unimodal spatial-discrimination task, the model includes occasional lapses at rate λ_{AV}. Therefore, the probability of reporting a visual stimulus at $s_{V,n}$ as located to the right of an auditory stimulus at location $s_{A,l}$ ($r_{l,n} = 1$) is equal to

$$p_{r_{l,n}=1} = 0.5\lambda_{AV} + (1-\lambda_{AV})p_{l,n}. \quad (31)$$

Pointing practice task. The pointing practice task was used to estimate localization response variability, σ_r^2, due to sources unrelated to the spatial perception of the stimuli. Visual stimuli were presented at eight different locations $s_{V,o}$, where o indexes the stimulus location ($o \in \{1, 2, \ldots, 8\}$). Localization responses (i.e., confirmed cursor positions) in each trial were modeled as perturbed by Gaussian-distributed noise and centered on the physical stimulus location:

$$r_{V,o} \sim \mathcal{N}(s_{V,o}, \sigma_r^2). \tag{32}$$

By doing so, we assume that 1) the location of the visual cursor in physical space maps directly to its location in perceptual space and 2) the stimulus location estimate is unbiased. This is based on our general assumption of identity remapping for visual stimuli as well as on the high spatial reliability of the visual cursor and the visual stimulus, which should safeguard their estimates against the influence of spatial priors. See Section S2 in S1 Appendix for a model that does not have these assumptions.

Unimodal localization task—Pre-recalibration phase. The unimodal localization task was conducted before and after the recalibration phase to measure shifts in auditory and visual localization responses as a consequence of exposure to spatially discrepant audiovisual stimuli during the recalibration phase. Stimuli were presented at four locations for each modality, $s_{A,l}$ and $s_{V,l}$.

As for the pointing practice task, we assumed that the probability distributions of the localization responses in physical space, $r_{A,l}$ and $r_{V_i,l}$, are centered on the location estimates $\hat{s}'_{A,l}$ and $\hat{s}'_{V_i,l}$ in perceptual space, and corrupted by additional unbiased noise. As in the spatial-discrimination tasks (and unlike in the pointing practice task that used a different, maximally reliable stimulus), the stimulus location estimates are assumed to be biased due to the remapping process and the incorporation of the supra-modal spatial prior. It follows that the probability distributions of the localization responses are

$$r_{A,l} \sim \mathcal{N}(\mu_{\hat{s}'_{A,l}}, \sigma^2_{\hat{s}'_A} + \sigma_r^2) \text{ and } r_{V_i,l} \sim \mathcal{N}(\mu_{\hat{s}'_{V_i,l}}, \sigma^2_{\hat{s}'_{V_i}} + \sigma_r^2), \tag{33}$$

where the terms are defined in Eqs 24–27.

Unimodal localization task—Post-recalibration phase. In the post-recalibration phase, the remapping from physical to perceptual space had been updated so that additional shifts Δ_{A_i} and Δ_{V_i}, accumulated after 120 exposures to discrepant audiovisual stimuli, were incorporated: $\hat{s}'_A = a_A s_A + b_A + \Delta_{A_i}(121)$ and $\hat{s}'_V = a_V s_V + b_V + \Delta_{V_i}(121)$. This change in the measurement distributions affects the centers of the location estimates' probability distributions as follows:

$$\begin{aligned} \mu_{\hat{s}'_{A_i,l}} &= \frac{\sigma'^{-2}_A (a_A s_{A,l} + b_A + \Delta_{A_i}(121)) + \sigma'^{-2}_P \mu'_P}{\sigma'^{-2}_A + \sigma'^{-2}_P} \\ \mu_{\hat{s}'_{V_i,l}} &= \frac{\sigma'^{-2}_{V_i} (s_{V,l} + \Delta_{V_i}(121)) + \sigma'^{-2}_P \mu'_P}{\sigma'^{-2}_{V_i} + \sigma'^{-2}_P}. \end{aligned} \tag{34}$$

We assumed that the probability distributions of the localization responses are centered on these updated values, that is, we did not implement the updated remapping for the location estimates of the cursor. In sum, localization responses to unimodally presented visual and auditory stimuli have a Gaussian probability distribution that, after the audiovisual recalibration task, additionally depends on the final shift updates $\Delta_{A_i}(121)$ and $\Delta_{V_i}(121)$.

Model fitting

All models were fit using a maximum-likelihood procedure. That is, a set of free parameters Θ was chosen to maximize the log likelihood of the data given a model M. Our model fitting strategy aimed to reduce the number of free parameters estimated at once. We split the set of free parameters into three subsets Θ_i, $i = 1, 2, 3$, each fit to a subset of the data X_i, $i = 1, 2, 3$, and maximized the log-likelihoods of each subset X_i separately,

$$\log\ P(X|M,\Theta) = \log\ P(X_1|M,\Theta_1) + \log\ P(X_2|M,\Theta_2) + \log\ P(X_3|M,\Theta_3), \tag{35}$$

where $X_1, X_2, X_3 \subset X$ and $\Theta_1, \Theta_2, \Theta_3 \subset \Theta$. The first dataset, X_1, refers to the unimodal spatial-discrimination task, which was used to constrain parameter subset, Θ_1, that comprised unimodal stimulus reliabilities as well as lapse rates in the different sessions of this task. X_2 refers to the pointing practice task used to estimate parameter subset, Θ_2, which comprised only the variability in localization responses due to other factors than spatial perception. The third subset comprised data from three tasks, $X_3^1, X_3^2, X_3^3 \subset X_3$, the bimodal spatial-discrimination task, and the unimodal spatial-localization task for the pre- and post-recalibration phase, respectively. These three datasets constrained overlapping sets of parameters $\Theta_3^1, \Theta_3^2, \Theta_3^3 \subset \Theta_3$. Thus, they were fit jointly. Θ_3^1 and Θ_3^2 comprised only localization bias parameters as well as task-specific lapse rates; stimulus reliabilities and response noise parameter estimates were taken from Θ_1 and Θ_2. Only Θ_3^3, the parameter set constrained by participants' post-recalibration localization responses, included parameters specific to each of the three models of cross-modal recalibration such as the learning rate and common-cause prior.

As outlined before, we did not fit the localization responses from the audiovisual recalibration task, i.e., the build-up of the recalibration effect (Section S13 in S1 Appendix), because the shifts ($\Delta_{A_i}(t)$ and $\Delta_{V_i}(t)$) are serially dependent, that is, the size of the shift in trial t depends on the size of the shift in trial $t - 1$. Given that there is no closed-form solution for the causal-inference model, we would have needed to use Monte Carlo simulations to approximate the probability distribution of the location estimates. Yet, the location estimates depend on the serially dependent shifts and consequently the number of necessary samples would have grown exponentially from trial to trial. Thus, it was computationally challenging to estimate the likelihood of the parameters and the model given the data from the audiovisual recalibration task. Instead, we used Monte Carlo simulations to approximate the probability distribution of the shift updates $\Delta_{A_i}(121)$ and $\Delta_{V_i}(121)$ accumulated at the end of the audiovisual recalibration task, i.e., we fitted the final recalibration effect rather than its build-up.

Model log-likelihood—Unimodal spatial-discrimination task. In the unimodal spatial-discrimination task (auditory session), participants indicated whether the test stimulus was located to the left, $r_{A,n(t)} = 0$, or to the right of the standard stimulus, $r_{A,n(t)} = 1$. For each such trial, the likelihood of model parameters given the response $r_{A,n(t)}$ is

$$P(r_{A,n(t)}|M,\Theta_1) = p_{r_{A,n(t)}=1}{}^{r_{A,n(t)}}\left(1 - p_{r_{A,n(t)}=1}\right)^{1-r_{A,n(t)}}, \tag{36}$$

where $p_{r_{A,n}=1}$ is defined in Eq 23. Thus, the log likelihood given responses across all $T_{1,A}$ trials is

$$\log\ P(X_{1,A}|M,\Theta_1) = \sum_{t=1}^{T_{1,A}} [r_{A,n(t)} \log\ p_{r_{A,n(t)}=1} + (1 - r_{A,n(t)}) \log(1 - p_{r_{A,n(t)}=1})]. \tag{37}$$

Analogously,

$$\log\ P(X_{1,V_i}|M,\Theta_1) = \sum_{t=1}^{T_{1,V_i}}[r_{V_i,n(t)}\log p_{r_{V_i,n(t)}=1} + (1-r_{V_i,n(t)})\log(1-p_{r_{V_i,n(t)}=1})]. \tag{38}$$

The log likelihood across all four sessions is

$$\log\ P(X_1|M,\Theta_1) = \log\ P(X_{1,A}|M,\Theta_1) + \sum_{i=1}^{3}\log\ P(X_{1,V_i}|M,\Theta_1). \tag{39}$$

The set of free parameters that were constrained by the binary responses in this task is $\Theta_1 = \{\sqrt{2}\sigma'_A/a_A, \sigma'_{V_1}, \sigma'_{V_2}, \sigma'_{V_3}, \lambda_A, \lambda_{V_1}, \lambda_{V_2}, \lambda_{V_3}\}$.

Model log-likelihood—Pointing practice task. For each trial, the likelihood of the model parameters given a visual stimulus at location $s_{V,o(t)}$ and a subsequent response (cursor setting) $r_{V,o(t)}$ is $P(r_{V,o(t)}|M,\Theta_2) = \varphi(r_{V,o(t)}; s_{V,o(t)}, \sigma_r^2)$ where φ refers to the Gaussian probability density. The only free parameter that was constrained by this task is $\Theta_2 = \{\sigma_r\}$. The maximum-likelihood estimate of σ_r is

$$\sigma_r = \sqrt{\frac{\sum_{t=1}^{T_2}(s_{V,o(t)} - r_{V,o(t)})^2}{T_2}}, \tag{40}$$

where T_2 and the sum do not include outlier trials.

Model log-likelihood—Bimodal spatial-discrimination task. In the bimodal spatial-discrimination task, for each trial t, participants indicated whether the visual test stimulus at location $s_{V_1,n(t)}$ was located to the left, $r_{l(t),n(t)} = 0$, or to the right, $r_{l(t),n(t)} = 1$, of the auditory standard stimulus presented at $s_{A,l(t)}$. For each such trial, the likelihood of model parameters given the response $r_{l(t),n(t)}$ is

$$P(r_{l(t),n(t)}|M,\Theta_3^1) = p_{r_{l(t),n(t)}=1}{}^{r_{l(t),n(t)}}(1-p_{r_{l(t),n(t)}=1})^{1-r_{l(t),n(t)}}, \tag{41}$$

where $p_{r_{l,n}=1}$ is defined in Eq 31. Thus, the log likelihood given the responses across all T_3^1 trials is

$$\log\ P(X_3^1|M,\Theta_3^1) = \sum_{t=1}^{T_3^1}[r_{l(t),n(t)}\log p_{r_{l(t),n(t)}=1} + (1-r_{l(t),n(t)})\log(1-p_{r_{l(t),n(t)}=1})]. \tag{42}$$

$p_{r_{l,n}=1}$ is a function of $p_{l(t),n(t)}$, which in turn depends on the bias parameters a_A and b_A, the parameters of the supra-modal prior over locations μ'_p and σ'_p, as well as the measurement variances σ'_A and σ'_{V_i} (see Eqs 24–27). Fitting both the bias parameters and the supra-modal prior at once was impossible as they effectively traded off. Thus, we implemented a non-informative supra-modal prior over stimulus locations by setting σ'_p to 100 and μ'_p to 0. $a\sigma'_A$, σ'_{V_1}, σ'_{V_2}, and σ'_{V_3} were estimated based on the forced-choice responses from the unimodal spatial-discrimination task. The final set of free parameters that were constrained by the binary responses in this task was $\Theta_3^1 = \{a_A, b_A, \lambda_{VA}\}$. The bias parameters, a_A and b_A, were jointly estimated using the data from this task as well as pre- and post-recalibration responses from the unimodal localization task.

Model log-likelihood—Unimodal localization task—Pre-recalibration phase. In this task, each localization results in cursor location settings $r_{A,i,j,l(t)}$ and $r_{V_i,j,l(t)}$ on trial t of session (i, j) where i indicates the visual-reliability condition and j the recalibration direction in the subsequent recalibration phase. The localization responses from this task were modeled as Gaussian-distributed. From these distributions, we can compute the likelihood of a model M

and the parameter set Θ_3^2 as the Gaussian probability density function in Eq 33 evaluated at the observed localization responses $r_{A,i,j,l(t)}$ and $r_{V_i,j,l(t)}$:

$$P(r_{A,i,j,l(t)}|M,\Theta_3^2) = (2\pi(\sigma^2_{\hat{s}'_A} + \sigma_r^2))^{-\frac{1}{2}} \exp\left[-\frac{(r_{A,i,j,l(t)} - \mu_{\hat{s}'_{A,i,j,l(t)}})^2}{2(\sigma^2_{\hat{s}'_A} + \sigma_r^2)}\right]$$
$$P(r_{V_i,j,l(t)}|M,\Theta_3^2) = (2\pi(\sigma^2_{\hat{s}'_{V_i}} + \sigma_r^2))^{-\frac{1}{2}} \exp\left[-\frac{(r_{V_i,j,l(t)} - \mu_{\hat{s}'_{V_i,j,l(t)}})^2}{2(\sigma^2_{\hat{s}'_{V_i}} + \sigma_r^2)}\right]. \quad (43)$$

The log likelihood is the sum of the log likelihoods across the trials of all six sessions:

$$\begin{aligned}
&\log P(X_3^2|M,\Theta_3^2) \\
&= \sum_{i=1}^{3}\sum_{j=1}^{2}\left[\sum_{t_A=1}^{T^2_{3,A,i,j}} \log P(r_{A,i,j,l(t_A)}|M,\Theta_3^2) + \sum_{t_{V_i}=1}^{T^2_{3,V_i,j}} \log P(r_{V_i,j,l(t_{V_i})}|M,\Theta_3^2)\right] \\
&= \sum_{i=1}^{3}\sum_{j=1}^{2}\Bigg[-\frac{T^2_{3,A,i,j}}{2}\log(2\pi(\sigma^2_{\hat{s}'_A} + \sigma_r^2)) - \frac{T^2_{3,V_i,j}}{2}\log(2\pi(\sigma^2_{\hat{s}'_{V_i}} + \sigma_r^2)) \\
&\quad - \frac{1}{2(\sigma^2_{\hat{s}'_A} + \sigma_r^2)}\sum_{t_A=1}^{T^2_{3,A,i,j}}(r_{A,i,j,l(t_A)} - \mu_{\hat{s}_{A,i,j,l(t_A)}})^2 \\
&\quad - \frac{1}{2(\sigma^2_{\hat{s}'_{V_i}} + \sigma_r^2)}\sum_{t_{V_i}=1}^{T^2_{3,V_i,j}}(r_{V_i,j,l(t_{V_i})} - \mu_{\hat{s}_{V_i,j,l(t_{V_i})}})^2\Bigg].
\end{aligned} \quad (44)$$

The log-likelihood depends on $\mu_{\hat{s}_{A,i,j,l(t_A)}}$ and $\mu_{\hat{s}_{V_i,j,l(t_{V_i})}}$, which in turn depend on the bias parameters a_A and b_A, the parameters of the supra-modal prior μ'_P and σ'_P, as well as the measurement variances σ'_A and σ'_{V_i} (see Eq 34), and the response noise σ_r. We chose a flat prior over stimulus locations, the (scaled) measurement variances ($\sqrt{2}\sigma'_A/a_A$ and σ'_{V_i}) were estimated based on the unimodal spatial-discrimination task, and σ_r was estimated based on the pointing practice task. Consequently, the actual set of parameters constrained by localization responses from the pre-recalibration task was $\Theta_3^2 = \{a_A, b_A\}$. Here, the values of the T variables and the sums do not include outlier trials.

Model log-likelihood—Unimodal localization task—Post-recalibration phase. Localization responses in the post-recalibration phase additionally depend on the updates for the visual and auditory shifts accumulated after 120 trials during the recalibration phase, $\Delta_{A_i}(121)$ and $\Delta_{V_i}(121)$ (Eq 34). Since these accumulated shift updates are not accessible to the experimenter, we marginalized over these shift updates to calculate the log-likelihood. For each of the six experimental sessions (i, j), the log likelihood of a model M and its parameter set Θ_3^3 is the integral over Δ_{A_i} and Δ_{V_i} of the likelihood of the final shift updates given the observed data $X^3_{3,i,j}$, the model M, and the parameter set Θ_3^3, $P(X^3_{3,i,j}|\Delta_{A_i,j}, \Delta_{V_i,j}, M, \Theta_3^3)$, multiplied by the joint probability of the auditory and visual shift updates, $P(\Delta_{A_i,j}, \Delta_{V_i,j}|M,\Theta_3^3)$, summed across all six sessions

$$\begin{aligned}
&\log P(X_3^3|M,\Theta_3^3) = \\
&\sum_{i=1}^{3}\sum_{j=1}^{2}\log\left(\int\int P(X^3_{3,i,j}|\Delta_{A_i,j}, \Delta_{V_i,j}, M, \Theta_3^3)P(\Delta_{A_i,j}, \Delta_{V_i,j}|M,\Theta_3^3)d\Delta_{A_i,j}d\Delta_{V_i,j}\right).
\end{aligned} \quad (45)$$

We will describe in the following sections how the joint probability $P(\Delta_{A_i j}, \Delta_{V_i j}|M, \Theta_3^3)$ and the log-likelihood $\log\ P(X_3^3|M, \Theta_3^3)$ were derived for each of the three models of cross-modal recalibration.

Reliability-based model of cross-modal recalibration. In this model, auditory and visual shift updates have a constant ratio of $\Delta_{A_i}(t)/\Delta_{V_i}(t) = -\sigma'^2_{AV,A}/\sigma'^2_{AV,V_i}$, the ratio of the measurement noise variances. Therefore, $\Delta_{A_i}(t)$ can be rewritten as $(-\sigma'^2_{AV,A}/\sigma'^2_{AV,V_i})\Delta_{V_i}(t)$, and we can express the likelihood given a single auditory localization response $r_{A_i j, l(t)}$ as

$$P(r_{A_i j,l(t)}|\Delta_{V_i j}, M_{RB}, \Theta^3_{3,RB}) = (2\pi(\sigma^2_{\hat{s}'_{A_i}} + \sigma_r^2))^{-\frac{1}{2}} \exp\left[-\frac{(r_{A_i j,l(t)} - \mu_{\hat{s}'_{A_i j,l(t)}})^2}{2(\sigma^2_{\hat{s}'_{A_i}} + \sigma_r^2)}\right], \tag{46}$$

where

$$\mu_{\hat{s}'_{A_i,l(t)}} = \frac{\sigma'^{-2}_{AV,A}\left(a_A s_{A,l(t)} + b_A - \frac{\sigma'^2_{AV,A}}{\sigma'^2_{AV,V_i}}\Delta_{V_i j}(121)\right) + \sigma'^{-2}_P \mu'_P}{\sigma'^{-2}_{AV,A} + \sigma'^{-2}_P}. \tag{47}$$

The visual response likelihoods and means are defined analogously (Eq 34). Thus, the joint likelihood $P(X^3_{3,i,j}|\Delta_{A_i j}, \Delta_{V_i j}, M_{RB}, \Theta^3_{3,RB})$ can be written as $P(X^3_{3,i,j}|\Delta_{V_i j}, M_{RB}, \Theta^3_{3,RB})$. Given that the likelihood depends only on $\Delta_{V_i j}$, we only need to integrate over $\Delta_{V_i j}$ and the log likelihood simplifies to

$$\log\ P(X_3^3|M_{RB}, \Theta^3_{3,RB}) = \sum_{i=1}^{3}\sum_{j=1}^{2} \log\left(\int P(X^3_{3,i,j}|\Delta_{V_i j}, M_{RB}, \Theta^3_{3,RB}) P(\Delta_{V_i j}|M_{RB}, \Theta^3_{3,RB}) d\Delta_{V_i j}\right). \tag{48}$$

The shift updates $\Delta_{V_i j}$ are stochastic because the visual and auditory measurements in each trial of the audiovisual recalibration task are stochastic. We cannot derive their probability distribution $P(\Delta_{V_i j}|M_{RB}, \Theta^3_{3,RB})$ in closed form. Instead, we used Monte Carlo simulation to approximate this probability distribution. Given the reliability-based model, for each candidate set of parameters $\Theta^3_{3,RB}$, visual-reliability condition *i*, and recalibration direction *j*, we simulated 120 recalibration trials analogous to the audiovisual recalibration task. We repeated this simulation 1,000 times, resulting in a sample of 1,000 shift updates (Δ_{V_i}) and checked whether the distribution of the 1,000 samples was well fit by a Gaussian with mean and standard deviation equal to the corresponding empirical parameters of the sampled distribution. To do so, we binned the simulated shift updates into 100 bins of equal size and computed the correlation between the observed and predicted number of samples per bin. The resulting value of R^2 was greater than 0.925 in all cases (Section S8 in S1 Appendix). The approximated probability distribution of the shift updates is denoted as $\tilde{P}(\Delta_{V_i j}|M_{RB}, \Theta^3_{3,RB})$.

We approximated the integral in Eq 48 by numerical integration over a region discretized into 100 bins. To ensure that we include enough of the tails of the probability distribution of the shift updates, we set the integration region to be three times larger than the range of the Δ_{V_i} samples, and centered the integration region on that range. Thus, the lower bound, *lb*, is defined as $lb = \Delta_{min} - (\Delta_{max} - \Delta_{min})$ and the upper bound is $ub = \Delta_{max} + (\Delta_{max} - \Delta_{min})$. The

numerical integration region was derived separately for each session. The log likelihood is:

$$\begin{aligned}
&\log P(X_3^3|M_{RB},\Theta_{3,RB}^3)\\
&=\sum_{i=1}^{3}\sum_{j=1}^{2}\log\left(\int P(X_{3,i,j}^3|\Delta_{V_i,j},M_{RB},\Theta_{3,RB}^3)P(\Delta_{V_i,j}|M_{RB},\Theta_{3,RB}^3)d\Delta_{V_i,j}\right)\\
&\approx\sum_{i=1}^{3}\sum_{j=1}^{2}\log\left(\int_{lb_{V_i,j}}^{ub_{V_i,j}}P(X_{3,i,j}^3|\Delta_{V_i,j},M_{RB},\Theta_{3,RB}^3)\tilde{P}(\Delta_{V_i,j}|M_{RB},\Theta_{3,RB}^3)d\Delta_{V_i,j}\right)\\
&\approx\sum_{i=1}^{3}\sum_{j=1}^{2}\log\left(\frac{ub_{V_i,j}-lb_{V_i,j}}{100}\sum_{k_V=1}^{100}P(X_{3,i,j}^3|\Delta_{V_i,j}(k_V),M_{RB},\Theta_{3,RB}^3\right)\times\\
&\quad\tilde{P}(\Delta_{V_i,j}(k_V)|M_{RB},\Theta_{3,RB}^3)),
\end{aligned}\tag{49}$$

where

$$\begin{aligned}
&P(X_{3,i,j}^3|\Delta_{V_i,j}(k_V),M_{RB},\Theta_{3,RB}^3)\\
&=\prod_{t_A=1}^{T_{3,A,i,j}^3}P(r_{A_i,j,l(t_A)}|\Delta_{V_i,j}(k_V),M_{RB},\Theta_{3,RB}^3)\prod_{t_{V_i}=1}^{T_{3,V_i,j}^3}P(r_{V_i,j,l(t_{V_i})}|\Delta_{V_i,j}(k_V),M_{RB},\Theta_{3,RB}^3)\\
&=(2\pi(\sigma_{\hat{s}_{A_i}}^2+\sigma_r^2))^{-\frac{T_{3,A,i,j}^3}{2}}\exp\left[-(2(\sigma_{\hat{s}_{A_i}}^2+\sigma_r^2))^{-1}\sum_{t_A=1}^{T_{3,A,i,j}^3}(r_{A_i,j,l(t_A)}-\mu_{\hat{s}'_{A_i,l(t_A)}})^2\right]\times\\
&\qquad(2\pi(\sigma_{\hat{s}_{V_i}}^2+\sigma_r^2))^{-\frac{T_{3,V_i,j}^3}{2}}\exp\left[-(2(\sigma_{\hat{s}_{V_i}}^2+\sigma_r^2))^{-1}\sum_{t_{V_i}=1}^{T_{3,V_i,j}^3}(r_{V_i,j,l(t_{V_i})}-\mu_{\hat{s}'_{V_i,l(t_{V_i})}})^2\right],
\end{aligned}\tag{50}$$

with $\mu_{\hat{s}'_{A,l(t)}}$ defined in Eq 47 and $\mu_{\hat{s}'_{V_i,l(t)}}$ defined in Eq 34. $\mu_{\hat{s}'_{A_i,l(t)}}$ and $\mu_{\hat{s}'_{V_i,l(t)}}$ depend on the bias parameters a_A and b_A, as well as on $\Delta_{V_i,j}$, which depends on the measurement variances $\sigma'_{AV,A}$ and σ'_{AV,V_i} given bimodal presentation and the common learning rate α. Note that $\sigma'_{AV,A}$ and σ'_{AV,V_i} are not directly constrained by data from bimodal trials (because these trials were not included in the model fitting), but estimated based on their influence on the shift updates. Specifically, $\sigma'_{AV,A}$ and σ'_{AV,V_i} affect the spread of the measurements, and as a consequence they influence the width of the predicted probability distribution of measurement-shift updates, which in turn affect the log likelihood of the model. The set of free parameters for this model is $\Theta_{3,RB}^3=\{a_A,b_A,\sigma'_{AV,A},\sigma'_{AV,V_1},\sigma'_{AV,V_2},\sigma'_{AV,V_3},\alpha\}$. σ'_{AV,V_i} was constrained to be a non-decreasing function of visual-reliability condition i, and $\sigma'_{AV,A}$ and σ'_{AV,V_i} were constrained to be no greater than five times the average values of σ'_A and σ'_{V_i} across participants (Section S14 in S1 Appendix). The values of the T variables and the sums do not include outlier trials.

Log-likelihood—fixed-ratio model of cross-modal recalibration. In this model, auditory and visual shift updates have a fixed ratio of $\Delta_{A_i}(t)/\Delta_{V_i}(t)=-\alpha_A/\alpha_V$ (Section S15 in S1 Appendix). Thus, we can express the likelihood for the fixed-ratio model and parameter set

$\Theta^3_{3,FR}$ given an auditory localization response $r_{A_i,j,l(t)}$ in a similar form to Eq 47:

$$\mu_{\hat{s}'_{A_i,l(t)}} = \frac{{\sigma'_{AV,A}}^{-2}\left(a_A s_{A,l(t)} + b_A - \frac{\alpha_A}{\alpha_V}\Delta_{V_i,j}(121)\right) + {\sigma'_P}^{-2}\mu'_P}{{\sigma'_{AV,A}}^{-2} + \sigma_P^{-2}}. \quad (51)$$

The approximation $\tilde{P}(\Delta_{V_i,j}|M_{FR}, \Theta^3_{3,FR})$ was generated in the same way as for the reliability-based model. The set of free parameters for this model is $\Theta^3_{3,FR} = \{a_A, b_A, \sigma'_{AV,A}, \sigma'_{AV,V_1}, \sigma'_{AV,V_2}, \sigma'_{AV,V_3}, \alpha_A, \alpha_V\}$. Note that even though the shift updates in the fixed-ratio model do not depend on the stimulus reliabilities, the log-likelihood does due to the influence of stimulus reliability on the estimates in the localization task (Eq 51) and due to the influence of the spread of the simulated measurements on the spread of the estimated distribution of $\tilde{P}(\Delta_{V_i,j}|M_{FR}, \Theta^3_{3,FR})$. As in the reliability-based model, σ'_{AV,V_i} was constrained to be a non-decreasing function of visual-reliability condition i, and $\sigma'_{AV,A}$ and σ'_{AV,V_i} were constrained to be no greater than five times the average values of σ'_A and σ'_{V_i} across participants.

Log-likelihood—causal-inference model of cross-modal recalibration. For this model, the joint likelihood $P(X^3_{3,i,j}|\Delta_{A_i,j}, \Delta_{V_i,j}, M_{CI}, \Theta^3_{3,CI})$ was truly two-dimensional. Thus, we approximated the joint probability of the auditory and visual shift updates, $P(\Delta_{A_i,j}, \Delta_{V_i,j}|M, \Theta^3_{3,CI})$ by drawing 1000 samples of shift-update pairs and compared the set of sample pairs to a 2-d Gaussian with the sample mean and covariance as parameters. We again tested whether the two-dimensional Gaussian distribution provided a good fit to the simulated density (defined as $R^2 > 0.925$). If the Gaussian fit was poor, we used a kernel density estimate (Gaussian kernel smoother with σ chosen automatically) of the distribution based on the 2-d density of the samples [77, 78]. Overall, the simulated auditory and visual shift updates were very well fit by a bivariate Gaussian, and we rarely used a kernel density estimate (Section S8 in S1 Appendix). We additionally used simulations to verify that our estimates of the partial model log-likelihood ($\log\ P(X^3_3|M_{CI}, \Theta^3_{3,CI})$) had reasonably small bias (Section S8 in S1 Appendix).

For the causal-inference model, we approximate the log likelihood by numerical integration over a 2-dimensional region of Δ_A, Δ_V space discretized into 100x100 bins. The upper and lower bounds were determined for both dimensions in the same way as before. The log likelihood is:

$$\begin{aligned}
&\log\ P(X^3_3|M_{CI}, \Theta^3_{3,CI}) \\
&\quad = \sum_{i=1}^{3}\sum_{j=1}^{2}\log\Bigg(\int\int P(X^3_{3,i,j}|\Delta_{A_i,j}, \Delta_{V_i,j}, M_{CI}, \Theta^3_{3,CI})\times \\
&\qquad\qquad P(\Delta_{A_i,j}, \Delta_{V_i,j}|M_{CI}, \Theta^3_{3,CI})d\Delta_{A_i,j}d\Delta_{V_i,j}\Bigg) \\
&\quad \approx \sum_{i=1}^{3}\sum_{j=1}^{2}\log\Bigg[\int_{lb_{V_i,j}}^{ub_{V_i,j}}\int_{lb_{A_i,j}}^{ub_{A_i,j}} P(X^3_{3,i,j}|\Delta_{A_i,j}, \Delta_{V_i,j}, M_{CI}, \Theta^3_{3,CI})\times \\
&\qquad\qquad \tilde{P}(\Delta_{A_i,j}, \Delta_{V_i,j}|M_{CI}, \Theta^3_{3,CI})d\Delta_{A_i,j}d\Delta_{V_i,j}\Bigg] \\
&\quad \approx \sum_{i=1}^{3}\sum_{j=1}^{2}\log\Bigg[\frac{ub_{V_i,j} - lb_{V_i,j}}{100}\frac{ub_{A_i,j} - lb_{A_i,j}}{100}\times \\
&\qquad\qquad \sum_{k_V=1}^{100}\sum_{k_A=1}^{100} P(X^3_{3,i,j}|\Delta_{A_i,j}(k_A), \Delta_{V_i,j}(k_V), M_{CI}, \Theta^3_{3,CI})\times \\
&\qquad\qquad\qquad P(\Delta_{A_i,j}(k_A), \Delta_{V_i,j}(k_V)|M_{CI}, \Theta^3_{3,CI})\Bigg],
\end{aligned} \quad (52)$$

Table 1. Summary of model parameters in Θ_3.

Θ	Meaning	RB	FR	CI	$CI_{\alpha_V=\alpha_A}$
$\sigma'_{AV,A}$	Measurement noise variance for the auditory stimulus in bimodal trials	✓	✓	✓	✓
σ'_{AV,V_i}	Measurement noise variance for the visual stimulus in bimodal trials	✓	✓	✓	✓
μ'_p	The mean of the supra- modal prior distribution	-	-	-	-
σ'_p	The standard deviation of the supra- modal prior	-	-	-	-
a_A, b_A	The slope and the intercept of the linear function that captures biases in auditory measurements	✓	✓	✓	✓
$P(C = 1)$	Prior probability of a common cause	-	-	✓	✓
α_A, α_V	Modality-specific learning rate	-	✓	✓	-
α	Common learning rate	✓	-	-	✓
λ_{AV}	Lapse rate for the bimodal spatial-discrimination task	✓	✓	✓	✓
	Number of total parameters:	8	9	10	9

https://doi.org/10.1371/journal.pcbi.1008877.t001

where $P(X^3_{3,i,j}|\Delta_{A_i,j}(k_A), \Delta_{V_i,j}(k_V), M_{CI}, \Theta^3_{3,CI})$ is defined analogously to the reliability-based model (see Eq 50) with the exception that $\mu_{\hat{s}'_{A,l(t)}}$ and $\mu_{\hat{s}'_{V_i,l(t)}}$ are defined in Eq 34. The set of free parameters used to fit the causal-inference model to the localization responses in the post-recalibration task is $\Theta^3_{3,CI} = \{a_A, b_A, \sigma'_{AV,A}, \sigma'_{AV,V_1}, \sigma'_{AV,V_2}, \sigma'_{AV,V_3}, \alpha_A, \alpha_V, p_{C=1}\}$ or $\Theta^3_{3,CI_{\alpha_V=\alpha_A}} = \{a_A, b_A, \sigma'_{AV,A}, \sigma'_{AV,V_1}, \sigma'_{AV,V_2}, \sigma'_{AV,V_3}, \alpha, p_{C=1}\}$.

Parameter estimation. For each model, we approximated the set of parameters Θ_1 and Θ_2 that maximized the likelihood using the MATLAB function fmincon and Python SciPy.optimize [79], and approximated Θ_3 using the BADS toolbox [80]. To deal with the possibility that the returned parameter values might correspond to a local minimum, we ran BADS multiple times with different starting points, randomly chosen from a D-dimensional grid, where D is the number of free parameters in Θ_3 (see Table 1 for a summary of the free parameters for each model) and with three evenly spaced values chosen for each dimension. The final parameter estimates were those with the maximum likelihood across all runs of the fitting procedure.

Model comparison. To compare model performance quantitatively, we computed the Akaike information criterion (AIC) for all four models [64] and calculated relative model-comparison scores, Δ_{AIC}, which relate the AIC value of the best-fit model to that of each of the other models (a higher Δ_{AIC} value indicates stronger evidence for the best-fit model). Models with $0 < \Delta_{\text{AIC}} < 2$ are weakly supported; models with $4 < \Delta_{\text{AIC}} < 7$ have considerably less support; models with $\Delta_{\text{AIC}} > 10$ have essentially no support [81].

Supporting information

S1 Appendix. Appendix. Supplemental control experiments, analyses and figures.
(PDF)

Acknowledgments

We would like to thank Luigi Acerbi for advice on model fitting, and Shannon Locke, Elyse Norton, Hörmet Yiltiz, Elon Gaffin-Cahn, Charlie Burlingham and Antonio Fernandez for support and comments. This work utilized the NYU IT High-Performance Computing resources and services.

Author Contributions

Conceptualization: Fangfang Hong, Stephanie Badde, Michael S. Landy.

Data curation: Fangfang Hong.

Formal analysis: Fangfang Hong.

Funding acquisition: Michael S. Landy.

Investigation: Fangfang Hong.

Methodology: Fangfang Hong, Stephanie Badde, Michael S. Landy.

Project administration: Michael S. Landy.

Resources: Michael S. Landy.

Software: Fangfang Hong, Stephanie Badde.

Supervision: Stephanie Badde, Michael S. Landy.

Validation: Fangfang Hong, Stephanie Badde, Michael S. Landy.

Visualization: Fangfang Hong.

Writing – original draft: Fangfang Hong, Stephanie Badde, Michael S. Landy.

Writing – review & editing: Fangfang Hong, Stephanie Badde, Michael S. Landy.

References

1. Mateeff S, Hohnsbein J, Noack T. Dynamic visual capture: apparent auditory motion induced by a moving visual target. Perception. 1985; 14:721–727. https://doi.org/10.1068/p140721 PMID: 3837873
2. Pick HL, Warren DH, Hay JC. Sensory conflict in judgments of spatial direction. Percept Psychophys. 1969; 6:203–205. https://doi.org/10.3758/BF03207017
3. Warren DH, Welch RB, McCarthy TJ. The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. Percept Psychophys. 1981; 30:557–564. https://doi.org/10.3758/BF03202010 PMID: 7335452
4. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Curr Biol. 2004; 14:257–262. https://doi.org/10.1016/j.cub.2004.01.029 PMID: 14761661
5. Battaglia PW, Jacobs RA, Aslin RN. Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am A. 2003; 20:1391–1397. https://doi.org/10.1364/JOSAA.20.001391 PMID: 12868643
6. Binda P, Bruno A, Burr DC, Morrone MC. Fusion of visual and auditory stimuli during saccades: a Bayesian explanation for perisaccadic distortions. J Neurosci. 2007; 27:8525–8532. https://doi.org/10.1523/JNEUROSCI.0737-07.2007 PMID: 17687030
7. Hartcher-O'Brien J, Di Luca M, Ernst MO. The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion. PLoS One. 2014; 9(3):e89339. https://doi.org/10.1371/journal.pone.0089339 PMID: 24594578
8. Raposo D, Sheppard JP, Schrater PR, Churchland AK. Multisensory decision-making in rats and humans. J Neurosci. 2012; 32:3726–3735. https://doi.org/10.1523/JNEUROSCI.4998-11.2012 PMID: 22423093
9. Sheppard JP, Raposo D, Churchland AK. Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. J Vis. 2013; 13(6):4. https://doi.org/10.1167/13.6.4 PMID: 23658374
10. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002; 415:429–433. https://doi.org/10.1038/415429a PMID: 11807554
11. Gepshtein S, Banks MS. Viewing geometry determines how vision and haptics combine in size perception. Curr Biol. 2003; 13:483–488. https://doi.org/10.1016/S0960-9822(03)00133-7 PMID: 12646130
12. Helbig HB, Ernst MO. Optimal integration of shape information from vision and touch. Exp Brain Res. 2007; 179:595–606. https://doi.org/10.1007/s00221-006-0814-y PMID: 17225091
13. Bresciani JP, Dammeier F, Ernst MO. Vision and touch are automatically integrated for the perception of sequences of events. J Vis. 2006; 6:554–564. https://doi.org/10.1167/6.5.2 PMID: 16881788

14. Angelaki DE, Gu Y, DeAngelis GC. Multisensory integration: psychophysics, neurophysiology, and computation. Curr Opin Neurobiol. 2009; 19:452–458. https://doi.org/10.1016/j.conb.2009.06.008 PMID: 19616425

15. Butler JS, Smith ST, Campos JL, Bülthoff HH. Bayesian integration of visual and vestibular signals for heading. J Vis. 2010; 10(11):23. https://doi.org/10.1167/10.11.23 PMID: 20884518

16. Fetsch CR, DeAngelis GC, Angelaki DE. Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur J Neurosci. 2010; 31:1721–1729. https://doi.org/10.1111/j.1460-9568.2010.07207.x PMID: 20584175

17. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. Nat Neurosci. 2011; 15:146–154. https://doi.org/10.1038/nn.2983 PMID: 22101645

18. Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. J Neurosci. 2009; 29:15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009 PMID: 20007484

19. Prsa M, Gale S, Blanke O. Self-motion leads to mandatory cue fusion across sensory modalities. J Neurophysiol. 2012; 108:2282–2291. https://doi.org/10.1152/jn.00439.2012 PMID: 22832567

20. Jack CE, Thurlow WR. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Percept Mot Skills. 1973; 37:967–979. https://doi.org/10.1177/003151257303700360 PMID: 4764534

21. Lewald J, Guski R. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Cogn Brain Res. 2003; 16:468–478. https://doi.org/10.1016/S0926-6410(03)00074-0 PMID: 12706226

22. Parise CV, Ernst MO. Correlation detection as a general mechanism for multisensory integration. Nat Commun. 2016; 7:11543. https://doi.org/10.1038/ncomms11543 PMID: 27265526

23. Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. Neuroreport. 2001; 12:7–10. https://doi.org/10.1097/00001756-200101220-00009 PMID: 11201094

24. Thurlow WR, Jack CE. Certain determinants of the "ventriloquism effect". Percept Mot Skills. 1973; 36:1171–1184. https://doi.org/10.2466/pms.1973.36.3c.1171 PMID: 4711968

25. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. Unifying multisensory signals across time and space. Exp Brain Res. 2004; 158:252–258. https://doi.org/10.1007/s00221-004-1899-9 PMID: 15112119

26. Acerbi L, Dokka K, Angelaki DE, Ma WJ. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. PLoS Comput Biol. 2018; 14(7):e1006110. https://doi.org/10.1371/journal.pcbi.1006110 PMID: 30052625

27. de Winkel KN, Katliar M, Bülthoff HH. Causal inference in multisensory heading estimation. PLoS One. 2017; 12(1):e0169676. https://doi.org/10.1371/journal.pone.0169676 PMID: 28060957

28. Gepshtein S, Burge J, Ernst MO, Banks MS. The combination of vision and touch depends on spatial proximity. J Vis. 2005; 5:1013–1023. https://doi.org/10.1167/5.11.7 PMID: 16441199

29. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. PLoS One. 2007; 2(9):e943. https://doi.org/10.1371/journal.pone.0000943 PMID: 17895984

30. Aller M, Noppeney U. To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. PLoS Biol. 2019; 17(4):e3000210. https://doi.org/10.1371/journal.pbio.3000210 PMID: 30939128

31. Badde S, Ley P, Rajendran SS, Shareef I, Kekunnaya R, Röder B. Sensory experience during early sensitive periods shapes cross-modal temporal biases. Elife. 2020; 9. https://doi.org/10.7554/eLife.61238

32. Badde S, Navarro KT, Landy MS. Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. Cognition. 2020; 197:104170. https://doi.org/10.1016/j.cognition.2019.104170 PMID: 32036027

33. Cao Y, Summerfield C, Park H, Giordano BL, Kayser C. Causal inference in the multisensory brain. Neuron. 2019; 102:1076–1087.e8. https://doi.org/10.1016/j.neuron.2019.03.043 PMID: 31047778

34. Dokka K, Park H, Jansen M, DeAngelis GC, Angelaki DE. Causal inference accounts for heading perception in the presence of object motion. Proc Natl Acad Sci U S A. 2019; 116:9060–9065. https://doi.org/10.1073/pnas.1820373116 PMID: 30996126

35. Locke SM, Landy MS. Temporal causal inference with stochastic audiovisual sequences. PLoS One. 2017; 12(9):e0183776. https://doi.org/10.1371/journal.pone.0183776 PMID: 28886035

36. McGovern DP, Roudaia E, Newell FN, Roach NW. Perceptual learning shapes multisensory causal inference via two distinct mechanisms. Sci Rep. 2016; 6:24673. https://doi.org/10.1038/srep24673 PMID: 27091411

37. Rohe T, Noppeney U. Sensory reliability shapes perceptual inference via two mechanisms. J Vis. 2015; 15(5):22:22. https://doi.org/10.1167/15.5.22 PMID: 26067540

38. Rohlf S, Li L, Bruns P, Röder B. Multisensory integration develops prior to crossmodal recalibration. Curr Biol. 2020; 30:1726–1732.e7. https://doi.org/10.1016/j.cub.2020.02.048 PMID: 32197090

39. Samad M, Chung AJ, Shams L. Perception of body ownership is driven by Bayesian sensory inference. PLoS One. 2015; 10(2):e0117178. https://doi.org/10.1371/journal.pone.0117178 PMID: 25658822

40. Wozny DR, Beierholm UR, Shams L. Probability matching as a computational strategy used in perception. PLoS Comput Biol. 2010; 6(8):e1000871. https://doi.org/10.1371/journal.pcbi.1000871 PMID: 20700493

41. Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. Nat Neurosci. 2006; 9:1432–1438. https://doi.org/10.1038/nn1790 PMID: 17057707

42. Bosen AK, Fleming JT, Brown SE, Allen PD, O'Neill WE, Paige GD. Comparison of congruence judgment and auditory localization tasks for assessing the spatial limits of visual capture. Biol Cybern. 2016; 110:455–471. https://doi.org/10.1007/s00422-016-0706-6 PMID: 27815630

43. Bruns P, Liebnau R, Röder B. Cross-modal training induces changes in spatial representations early in the auditory processing pathway. Psychol Sci. 2011; 22:1120–1126. https://doi.org/10.1177/0956797611416254 PMID: 21771962

44. Frissen I, Vroomen J, de Gelder B, Bertelson P. The aftereffects of ventriloquism: are they sound-frequency specific? Acta Psychol. 2003; 113:315–327. https://doi.org/10.1016/S0001-6918(03)00043-X PMID: 12835002

45. Frissen I, Vroomen J, de Gelder B, Bertelson P. The aftereffects of ventriloquism: generalization across sound-frequencies. Acta Psychol. 2005; 118:93–100. https://doi.org/10.1016/j.actpsy.2004.10.004 PMID: 15627411

46. Kopčo N, Lin IF, Shinn-Cunningham BG, Groh JM. Reference frame of the ventriloquism aftereffect. J Neurosci. 2009; 29:13809–13814. https://doi.org/10.1523/JNEUROSCI.2783-09.2009 PMID: 19889992

47. Lewald J. Rapid adaptation to auditory-visual spatial disparity. Learn Mem. 2002; 9:268–278. https://doi.org/10.1101/lm.51402 PMID: 12359836

48. Radeau M, Bertelson P. The after-effects of ventriloquism. Q J Exp Psychol. 1974; 26:63–71. https://doi.org/10.1080/14640747408400388 PMID: 4814864

49. Recanzone GH. Rapidly induced auditory plasticity: the ventriloquism aftereffect. Proc Natl Acad Sci USA. 1998; 95:869–875. https://doi.org/10.1073/pnas.95.3.869 PMID: 9448253

50. Watson DM, Akeroyd MA, Roach NW, Webb BS. Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history. Sci Rep. 2019; 9(1):8513. https://doi.org/10.1038/s41598-019-44984-9 PMID: 31186503

51. Wozny DR, Shams L. Recalibration of auditory space following milliseconds of cross-modal discrepancy. J Neurosci. 2011; 31:4607–4612. https://doi.org/10.1523/JNEUROSCI.6079-10.2011 PMID: 21430160

52. Canon LK. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. J Exp Psychol. 1970; 84:141–147. https://doi.org/10.1037/h0028925 PMID: 5480918

53. Ghahramani Z, Wolpert DM, Jordan MI. Computational models of sensorimotor integration. In: Morasso P, Sanguineti V, editors. Self-Organization, Computational Maps, and Motor Control. vol. 119 of Advances in Psychology. New York: Elsevier; 1997. p. 117–147.

54. Welch RB, Warren DH. Immediate perceptual response to intersensory discrepancy. Psychol Bull. 1980; 88:638–667. https://doi.org/10.1037/0033-2909.88.3.638 PMID: 7003641

55. Atkins JE, Jacobs RA, Knill DC. Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. Vision Res. 2003; 43:2603–2613. https://doi.org/10.1016/S0042-6989(03)00470-X PMID: 14552802

56. Di Luca M, Machulla TK, Ernst MO. Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. J Vis. 2009; 9(12):7. https://doi.org/10.1167/9.12.7 PMID: 20053098

57. Ernst MO, Banks MS, Bülthoff HH. Touch can change visual slant perception. Nat Neurosci. 2000; 3:69–73. https://doi.org/10.1038/71140 PMID: 10607397

58. Fujisaki W, Shimojo S, Kashino M, Nishida S. Recalibration of audiovisual simultaneity. Nat Neurosci. 2004; 7:773–778. https://doi.org/10.1038/nn1268 PMID: 15195098

59. van Beers RJ, Wolpert DM, Haggard P. When feeling is more important than seeing in sensorimotor adaptation. Curr Biol. 2002; 12:834–837. https://doi.org/10.1016/S0960-9822(02)00836-9 PMID: 12015120

60. Vroomen J, Keetels M, de Gelder B, Bertelson P. Recalibration of temporal order perception by exposure to audio-visual asynchrony. Cogn Brain Res. 2004; 22:32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003 PMID: 15561498

61. Burge J, Girshick AR, Banks MS. Visual-haptic adaptation is determined by relative reliability. J Neurosci. 2010; 30:7714–7721. https://doi.org/10.1523/JNEUROSCI.6427-09.2010 PMID: 20519546

62. Zaidel A, Turner AH, Angelaki DE. Multisensory calibration is independent of cue reliability. J Neurosci. 2011; 31:13949–13962. https://doi.org/10.1523/JNEUROSCI.2732-11.2011 PMID: 21957256

63. Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput. 2007; 19:3335–3355. https://doi.org/10.1162/neco.2007.19.12.3335 PMID: 17970656

64. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics, Vol.I, Foundations and Basic Theory. New York: Springer; 1992. p. 610–624.

65. Ernst M, Di Luca M. Multisensory perception: From integration to remapping. In: Trommershäuser J, Körding K, Landy MS, editors. Sensory Cue Integration. New York: Springer; 2011. p. 224–250.

66. Odegaard B, Wozny DR, Shams L. Biases in visual, auditory, and audiovisual perception of space. PLoS Comput Biol. 2015; 11(12):e1004649. https://doi.org/10.1371/journal.pcbi.1004649 PMID: 26646312

67. Bosen AK, Fleming JT, Allen PD, O'Neill WE, Paige GD. Accumulation and decay of visual capture and the ventriloquism aftereffect caused by brief audio-visual disparities. Exp Brain Res. 2017; 235:585–595. https://doi.org/10.1007/s00221-016-4820-4 PMID: 27837258

68. Doucet A, De Freitas N, Gordon NJ. Sequential Monte Carlo methods in practice. vol. 1(2). Springer; 2001.

69. Vercillo T, Gori M. Attention to sound improves auditory reliability in audio-tactile spatial optimal integration. Front Integr Neurosci. 2015; 9:34. https://doi.org/10.3389/fnint.2015.00034 PMID: 25999825

70. Frassinetti F, Bolognini N, Ladavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res. 2002; 147:332–343. https://doi.org/10.1007/s00221-002-1262-y PMID: 12428141

71. Passamonti C, Frissen I, Ladavas E. Visual recalibration of auditory spatial perception: two separate neural circuits for perceptual learning. Eur J Neurosci. 2009; 30:1141–1150. https://doi.org/10.1111/j.1460-9568.2009.06910.x PMID: 19735289

72. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10:433–436. https://doi.org/10.1163/156856897X00357 PMID: 9176952

73. Kleiner M, Brainard D, Pelli D, Ingling A, Murray RF, Broussard C. What's new in Psychtoolbox-3? Perception. 2007; 36:1–16.

74. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis. 1997; 10:437–442. https://doi.org/10.1163/156856897X00366 PMID: 9176953

75. Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971; 49:467–477. https://doi.org/10.1121/1.1912375 PMID: 5541744

76. Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys. 2001; 63:1293–1313. https://doi.org/10.3758/BF03194544 PMID: 11800458

77. Silverman BW. Density estimation for statistics and data analysis. vol. 26. London: CRC press; 1986.

78. Bowman AW, Azzalini A. Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations. vol. 18. New York: Oxford University Press; 1997.

79. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020; 17:261–272. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

80. Acerbi L, Ma WJ. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems. vol. 20. San Francisco: Curran Associates; 2017. p. 1834–1844.

81. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Ed. New York: Springer; 2002.