

Seeing Beyond Sight: Multisensory Inference under Degraded Visual Input

Thesis by
Yeuk Chin Ailene CHAN

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended May 12th, 2026

© 2026

Yeuk Chin Ailene CHAN
ORCID: 0000-0001-9522-0417

All rights reserved

ACKNOWLEDGEMENTS

Getting a PhD is not an easy journey. There were moments when it felt impossible. But if I had a time machine and could return to September 2021, I would still pack my life into three suitcases, cross the Pacific Ocean, and start anew in a foreign place.

It takes a village to raise a child—and perhaps an even bigger village to shape a PhD student. I am deeply grateful to all the people who supported, challenged, and guided me along the way.

First and foremost, I would like to thank my advisor, Prof. Shinsuke Shimojo, for showing genuine enthusiasm and passion for science from the very first day. Your approach to research has profoundly influenced the way I think and work. I am also grateful to my committee members, Prof. John O’Doherty, Prof. Ueli Rutishauser, and Prof. Ladan Shams, for their thoughtful discussions and constructive feedback.

Thank you to the SLab members, especially Dr. Daw-An Wu, for being both a scientific mentor and a steady source of life advice. I also want to thank my collaborators, Dr. Noelle Stiles, Dr. Carmel Levitan, and Prof. Armand Tanguay, for bringing in fresh perspectives to my projects. A special thanks to Dr. Daniel Wagenaar for his invaluable support during my work with the vibrotactile belt. There were more technical challenges than we’d like to admit, but I am proud of our ability to think on our feet and brave through the data collection phase.

Beyond my mentors at Caltech, I am deeply indebted to Professor Dorita Chang, my first scientific mentor. Thank you for taking a chance on a clueless undergraduate and for training me to be an independent researcher. I learned so much from you, both within the lab and beyond.

I would like to express my deepest gratitude to the Croucher Foundation for their unwavering support and belief in me. This PhD would not have been possible without their generosity. When the funding situation was dire, they stepped in and provided an additional year of support—an act of kindness and generosity that I will always remember.

Being far from home and venturing into unknown scientific territories came with many moments of feeling lost. During those times, I found grounding by reconnecting with the hobbies I loved as a child. I am grateful to Caltech Chamber Music for

giving me the chance to rediscover the joy of making music. I would also like to thank the Pasadena Dance Theatre and the wonderful group of ladies I dance with. They welcomed me with open arms, and it has been genuinely inspiring to see many of them dancing regularly for decades. And my ballet teacher is turning 100 this year—can you believe that?

As someone who is not naturally social, I feel incredibly lucky to have found my community here at Caltech. Thank you for bringing laughter into long weeks, for supporting each other when science gets a little too difficult, for celebrating both the big and small wins together, and for sharing our love and passion for culinary delights.

I would like to thank my husband, Ryan, for being by my side and for believing in me when I struggle to. You are the most wonderful person I've ever met. Thank you for your unparalleled ability to make me smile and soothe my racing mind. I truly could not have done this without you. I am so excited to build our life together, and I cannot wait for our wedding.

Above all, I am immensely grateful to my parents. I know how difficult it must have been to watch your only child leave home and pursue a PhD halfway across the world. Thank you for having faith in me and supporting my decisions. I carry your love with me in everything I do, and I hope with all my heart that I make you proud. 爸爸媽媽，女兒要畢業了，我長大成家了，不用再擔心我了！

Lastly, I want to thank myself. You have come a long way from the little girl who felt terrified when the path she had trained for her whole life no longer felt right. You worked extremely hard, never once took a break, and arrived at where you once only dreamed of being. This five-year journey has taught you far more than science or research. You learned about resilience, vulnerability, and the importance of asking for support. Thank you for reaching out when you needed company, for seeking help when you were overwhelmed, and for finally standing up for yourself. You have grown in ways that the younger version of you could never have imagined.

And a little shoutout to my cat, Panini. Thank you for being so perfect and sitting with me through the long writing days.

ABSTRACT

Navigating a complex environment requires the human nervous system to continuously integrate noisy and often conflicting sensory signals. To do so, the brain must solve the causal inference problem: determining whether signals arise from a common source and should be integrated, or from independent sources and should remain segregated. While Bayesian Causal Inference (BCI) provides a principled framework for this computation, it remains unclear how these mechanisms operate under conditions of severely degraded or absent visual input.

This dissertation investigates multisensory inference across a spectrum of visual uncertainty using complementary behavioral and computational approaches in three regimes: clinical low vision, physiological absence at the blind spot, and controlled degradation in immersive virtual reality with visuotactile interactions. In the low vision study, we find that multisensory illusions are not enhanced despite chronic impairment. Instead, BCI accurately captures behavior at high-visibility locations, where inference parameters remain comparable to sighted controls. At low-visibility locations, model fits become substantially weaker and behavioral responses more variable, limiting reliable interpretation of fitted parameters. Importantly, this likely reflects increased behavioral noise and the exploratory nature of the experimental design, which was optimized for broad spatial sampling rather than exhaustive inclusion of conditions required for robust model fitting at each location.

In the blind spot, where visual input is structurally absent, illusory percepts persist and remain highly structured. BCI provides an excellent account of behavior, revealing stable binding tendencies alongside increased visual uncertainty and broader priors. This demonstrates that multisensory filling-in arises from probabilistic inference under uncertainty rather than strengthened multisensory binding.

In a virtual reality visuotactile task, visual reliability is systematically manipulated in a naturalistic 3D environment. Here, BCI continues to explain behavior best under high reliability, but its predictive power decreases as visual uncertainty increases. We see a similar pattern of increased visual uncertainty and broader priors in response to experimentally reduced visual reliability. Notably, BCI model fit scales with task performance, with higher accuracy observed in individuals whose behavior more closely approximates a Bayes-optimal observer, indicating more effective multisensory integration.

Across all three studies, the results reveal a consistent pattern: the brain's causal inference architecture remains stable across vastly different visual conditions. Rather than increasing binding tendency, the system adapts by recalibrating sensory uncertainty and prior expectations. However, as uncertainty increases, behavior deviates from Bayes-optimal predictions, reflecting both computational and representational limitations.

Together, this work refines our understanding of sensory compensation by showing that multisensory interactions are shaped primarily by local sensory reliability rather than broad changes in binding tendency. By identifying when and why BCI succeeds or falls short, this dissertation provides a more nuanced account of perception under degraded conditions and offers useful considerations for the design of sensory substitution systems.

PUBLISHED CONTENT AND CONTRIBUTIONS

- [1] Ailene Y. C. Chan, Noelle R. B. Stiles, Carmel A. Levitan, Armand R. Tanguay, and Shinsuke Shimojo. Did you see the sound? a bayesian perspective on crossmodal perception in low vision. *bioRxiv*, 2025. doi: 10.64898/2025.12.24.696433.

A.Y.C.C. participated in the conception of the project, designed and performed the experiment, collected and analyzed the data, and participated in the writing of the manuscript.

- [2] Ailene Y. C. Chan, Noelle R. B. Stiles, Carmel A. Levitan, Daw-An Wu, Armand R. Tanguay, and Shinsuke Shimojo. Bayesian causal inference accounts for multisensory filling-in at the blind spot. *bioRxiv*, 2025. doi: 10.1101/2024.11.15.623713.

A.Y.C.C. conceived the project, designed and performed the experiment, collected and analyzed the data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Published Content and Contributions	vii
Table of Contents	vii
List of Illustrations	ix
List of Tables	xxiii
Chapter I: Introduction	1
Chapter II: Theoretical Framework	4
Chapter III: Multisensory inference in low vision	11
3.1 Introduction	12
3.2 Methods	13
3.3 Results	21
3.4 Discussion	30
3.5 Conclusion	34
Chapter IV: Multisensory inference at the blind spot	36
4.1 Introduction	36
4.2 Methods	38
4.3 Results	45
4.4 Discussion	52
4.5 Conclusion	55
Chapter V: Multisensory inference in virtual environment	57
5.1 Introduction	57
5.2 Methods	59
5.3 Results	70
5.4 Discussion	83
5.5 Conclusion	86
Chapter VI: General Conclusion	87
Bibliography	91
Appendix A: Supplemental Information for Chapter III	102
Appendix B: Supplemental Information for Chapter IV	110
Appendix C: Supplemental Information for Chapter V	117

LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
2.1 Visualization of MLE Model. Left: Model behavior when $\sigma_V = \sigma_T$. Right: Model behavior when $\sigma_V < \sigma_T$	5
2.2 Visualization of BCI Model. Left: Model behavior when the common cause ($C = 1$) hypothesis is assumed. Right: Model behavior when the independent cause ($C = 2$) hypothesis is assumed.	6
2.3 Simulated visual estimates under varying sensory reliability and spatial priors. Each panel displays the spatial prior ($p(s)$, peach), the visual likelihood (x_V , purple), and the tactile likelihood (x_T , green). The dashed vertical line indicates the predicted visual estimate ($\hat{s}_V$) under the BCI model using model averaging, FF, and FS models. Posterior probability of a common cause $P(C = 1 x_V, x_T)$ is fixed at 0.6. Top Row: Equal reliability ($\sigma_V = \sigma_T = 2.0$). Bottom Row: Tactile uncertainty exceeds visual uncertainty ($\sigma_V = 1.0, \sigma_T = 3.0$), simulating conditions of high visual reliability. Columns (Left to Right): The spatial prior becomes progressively flatter ($\sigma_P = 1.5, 3, 6$), demonstrating a decreasing “central bias” and increasing reliance on sensory data.	9
3.1 Stimuli and conditions (A) Stimulus locations corresponding to eight equally spaced positions along the circumference at 5° , 10° , and 15° eccentricities from central fixation. (B-E) Schematic of the four main experimental conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps). When paired, visual flashes (16.7 ms) and auditory beeps (7 ms) had simultaneous onsets.	16
3.2 Mean number of flashes perceived by task across all stimulus locations. Top row: <i>Visual Flash Detection task</i> (F1B0 condition). Bottom row: <i>Illusory Double Flash task</i> (F1B2 condition). Left panel: Low-vision group. Right panel: Sighted control group. Stimulus locations were further categorized for each participant as high-visibility (flash detection accuracy $> 50\%$) or low-visibility ($\leq 50\%$) based on performance in the detection task.	21

3.3	Mean accuracy of flash perception across conditions. (A) Heatmaps illustrate the mean accuracy in reporting the veridical number of flashes (0 or 1) across experimental conditions. Locations were categorized into high-visibility (flash-detection accuracy > 50%) and low-visibility ($\leq 50\%$) cohorts based on individual performance in the Visual Flash Detection Task (top panels). Bottom panels display accuracy as a function of stimulus eccentricity (5° , 10° , and 15°). Notably, in the F1B2 (1 flash, 2 beeps) condition, higher accuracy reflects a reduction in the sound-induced flash illusion. Inaccuracy among Low vision participants was split between “misses” (0-flash) and “illusions” (2-flashes), whereas inaccuracy among sighted control participants was driven by a shift toward reporting ‘illusions’ (2-flashes). F0B0 trials yielded the highest overall accuracy, reflecting reliable reports of zero flashes in the absence of stimulation. (B) Proportion of stimulus locations classified as high-visibility for each participant. (C) Spatial layout of stimulus locations, color-coded by eccentricity (5° = pink, 10° = purple, 15° = green), provided for illustrative purposes.	23
3.4	Correlation between flash detection accuracy and perceived flash counts. Scatter plots show the relationship between flash-detection accuracy (F1B0; Visual Flash Detection Task) and the average number of flashes perceived in the one-beep (F1B1, left) and two-beep (F1B2, right) conditions in the Illusory Double Flash Task. Each data point represents a participant’s detection accuracy and their perceived flash count at a specific stimulus location and eye. (48 data points per participant: 24 locations $\times$ 2 eyes) In the F1B1 condition (left), data points represent discrete judgments from individual catch trials presented at each location. In the F1B2 condition (right), data points represent the mean perceived flash count across ten repetitions. Top row: Low Vision; Bottom row: Sighted Control.	25

- 3.5 Model comparison and selection across stimulus location subsets.** BIC evaluates model performance as a function of free parameters and predictive power, lower score means better model performance. **(A)** Raw BIC scores for the four candidate models (BCI, MLE, FF, and FS) across all spatial partitions for Control (left) and low vision (right) groups. Lower BIC scores indicate superior model performance, balancing goodness-of-fit with model parsimony. **(B)** Total trial counts (n) per stimulus location subset, illustrating the data volume used for each model fit. **(C)** Normalized BIC scores (BIC/n), accounting for the varying trial counts across subsets. This normalization reveals that while raw BIC scores may appear lower for low-visibility locations due to smaller sample sizes, the predictive power of the Bayesian framework is optimized at high-visibility locations. Note the consistent superiority of the BCI model (lowest BIC/n) across both groups in high-visibility conditions. 27
- 3.6 Group-level model fits and parameter distributions for high- and low-visibility locations.** Mean response probabilities (solid lines) and BCI model predictions (dashed lines) for low vision and sighted control groups across stimulus conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps). 28

- 4.1 **Experimental Design** (A) Schematic of the six experimental conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps). When paired, visual flashes (16.7 ms) and auditory beeps (7 ms) had simultaneous onsets. Key conditions include the **Illusory AV Rabbit (F2B3)**, where an illusory flash is perceived with the second beep, and the **Invisible AV Rabbit (F3B2)**, where the middle flash is perceptually suppressed by the missing second beep. (B) Spatial configuration of stimuli shown for the left eye. Flashes were presented at the center of the blind spot (grey area) or a corresponding control spot (hollow area), as well as 0.5 visual degrees from the cardinal borders of these spots. In *F2BY* conditions, the two flashes appeared at the solid border points. In *F3BY* conditions, the first and third flashes appeared at the borders, while the second flash was presented at the center of the blind spot or control spot. For illustration, flashes traveling across the blind spot and control spot are colored blue and red, respectively; all flashes in the experiment were white. (C) The response interface for AV Rabbit Task. Participants reported the number of flashes they perceived by selecting from five equidistant options. Multiple selections were permitted. The colors are for illustrative purposes only. The response interface for the *Additional Flash Localization Task* is shown in Supplementary Video 1. (D) Overlay of the mapped blind spots across participants ($N = 20$). Blind-spot locations and sizes were consistent across individuals (see Fig. S1). 40

- 4.2 Effect of beeps on flash perception.** (A) Accuracy for reporting flashes (left) and beeps (right) across control and across-blind-spot stimulus locations. Flash accuracy decreased markedly in conditions where the number of flashes and beeps mismatched (F2B3 and F3B2). Participants also showed reduced accuracy for three-flash stimuli presented across the blind spot in the unimodal (F3B0) condition. Beep-report accuracy remained near ceiling across all conditions. (B) Illusion rates in the Illusory Audiovisual Rabbit (F2B3) and Invisible Audiovisual Rabbit (F3B2) conditions were comparable between stimulus locations. (C) Reported flash positions in the illusion conditions. In F2B3, the extra beep induced an illusory flash between the two real flashes; in F3B2, the missing beep suppressed perception of the middle flash between the first and third flashes. (D) Signal detection results for reporting three flashes. Sensitivity (d') was higher at control than across-blind-spot locations. Response criterion (c) was similar between locations by visual inspection, but it was consistently more conservative across the blind spot, indicating a greater tendency to report “2” across the blind spot. (E) Two-dimensional kernel density maps showing spatial offsets (in degrees of visual angle) of perceived illusory flashes (left) and last-flash reports (right). Offsets are computed relative to the center of the blind-spot/control region for illusory flashes, and relative to the true physical location of the last flash for last-flash localization errors. Horizontal and vertical offsets represent signed deviations along each axis. 48

- 4.3 The generative model of the Bayesian Causal Inference model and group-average model fitting results.** (A) The BCI model where multisensory cues either share a common cause ($C=1$) or independent causes ($C=2$). If cues share a common cause, the final percept s is derived by cues from both modalities, i.e., x_V and x_A , each perturbed by noises with width σ_V and σ_A , respectively. If cues are independent, final percepts s_V and s_A are derived independently from x_V and x_A . (B) The BCI model infers the causal structure by combining sensory likelihood and prior. Prior parameters include prior stimulus expectation and p_{common} (*a priori* expectation of a common cause). Sensory likelihood represents the sensory input (x_V, x_A) perturbed noise (σ_V, σ_A). Here, we have the group-average fitted parameters across-blind-spot and the predicted probability of auditory and visual responses. Since auditory likelihood is narrow compared to visual likelihood, the final percept s is heavily influenced by the auditory stimulus. (C) Group average model fitting results. Solid purple and green lines correspond to the visual and auditory response data, dashed dark purple and dark green lines correspond to the model predictions for visual and auditory responses. Given the robust BCI fits (mean $R^2 = 0.99$), we see almost a complete overlap of response data and model predictions. 51
- 5.1 Vibrotactile Belt Prototype and Control Hardware.** (A) The wearable belt assembly featuring eight tactors distributed at cardinal and collateral positions. (B) The custom PCB interface integrated with an Arduino Nano 33 BLE Sense Rev2 for signal processing and wireless communication. 60
- 5.2 Visual assets and integrated environment.** (A) Base terrain showing surface texture and starfield background. (B) Unidentified Flying Objects (UFO) model. (C) Laser gun used for player interaction. (D) Final composite scene of the main menu interface demonstrating the integration of all assets within the game engine. (E) Calibration scene. 61
- 5.3 Visual and Tactile Stimulus Directions** A: Stimulus directions categorized into three regions: front, side, and back. B-D: Visual (purple circle) and possible tactile (green triangle) stimuli directions in VT-con, VT-small, and VT-large conditions 62

5.4	Visual Reliability Levels. (A) High Reliability ($\alpha = 0$). (B) Low Reliability ($\alpha = 245$).	63
5.5	Group Performance ($N = 25$): Accuracy and Standard Deviation (SD) Across Visual Reliability Conditions. Line plots represent mean accuracy ($[0, 1]$) across conditions; box plots illustrate the distribution of response standard deviation (SD). For visual responses (participants shoot at UFO), dark purple denotes high visual reliability trials, while light purple denotes low visual reliability. For tactile responses (participants shoot at direction of vibration), brown and yellow represent high and low visual reliability, respectively. Green indicates unimodal tactile control trials (no visual stimuli). See Fig. C.2 for mean accuracy pooled across all directions.	71
5.6	Mean Directional Bias and Response Precision across Spatial Disparity and Visual Reliability Polar plots illustrate the aggregate response vectors for visual responses. The angular position of each arrow represents the mean directional bias ($\bar{\theta}$), where 0° (horizontal) denotes the visual target and the gray shaded region represents the conflict zone toward the tactile target. The length of each arrow represents the vector strength (R), a measure of circular precision ranging from 0 (random) to 1 (perfect consensus).	73
5.7	Group Average Movement Trajectories across Conditions and Visual Reliability Levels. Angular distance from the target ($^\circ$) is plotted relative to shot time (0s) for both Heading Direction (Left) and Controller Direction (Right). Solid lines indicate the group mean ($N = 25$), with shaded regions representing ± 1 SEM. Vertical dashed lines indicate the mean reaction time for each respective condition and reliability level. Negative time values represent the approach phase preceding the final shot event.	75

- 5.8 **Random Forest Classification Performance and Kinematic Feature Importance.** Each row represents a distinct classification task performed on movement kinematics ($N = 25$). **Left column:** Normalized confusion matrices illustrating the proportion of trials classified for each class label; dark diagonals indicate high classification accuracy. **Right column:** Top 5 predictors for each model ranked by Gini Importance, which quantifies a feature's contribution to the model. Feature labels indicate whether the data was recorded from the *Head* or *Controller* sensor. 77
- 5.9 **Model Comparison by Visual Reliability.** Mean Bayesian Information Criterion (BIC) scores (bar plots, left axis; lower is better) and mean R^2_{MF} (line plots, right axis; higher is better) for the four models tested: BCI, MLE, FS, FF. Error bars represent ± 1 standard error of the mean (SEM). Asterisks denote statistical significance (n.s. Not Significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$). 79
- 5.10 **Correlation between behavioral accuracy and BIC scores (Model performance; the lower the better), and R^2_{MF} values (Goodness of fit; the higher the better).** Each data point shows a participant's mean visual/tactile accuracy versus their BIC score and R^2_{MF} values from BCI and MLE model fitting. Blue trend lines represent BCI model, orange trend lines represent MLE model. 80
- 5.11 **Group average Bayesian Causal Inference model fits for Front region.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° 82
- A.1 **Mean Accuracy of Perceiving Beeps (N=16).** Heatmaps show accuracy in reporting the true number of beeps (0/1/2) across conditions. Performance is near ceiling (1.0) across conditions, suggesting that number of flashes present has little influence on beep detection accuracy. 103

- A.2 Sensitivity Analysis for different σ_A values.** To validate the selection of the fixed auditory likelihood parameter (σ_A), we performed a parameter sweep across a range of values (0.1, 0.2, 0.3, 0.4, 0.6, 0.8). By visual inspection, it was apparent that 0.1-0.3 yielded the best model performance (low BIC scores and high R^2). We performed Wilcoxon signed-rank tests to compare BIC and R^2 between 0.1 and 0.2, and 0.2 and 0.3. After correcting for multiple comparisons (Bonferroni), $\text{Model}_{\sigma_A=0.3}$ performs significantly worse than $\text{Model}_{\sigma_A=0.2}$. Among the control group, while $\text{Model}_{\sigma_A=0.1}$ has significantly lower BIC scores than $\text{Model}_{\sigma_A=0.2}$, there is no meaningful difference in R^2 . 104
- A.3 Auditory influence on flash perception.** Both groups show auditory influence on the number of flashes perceived. Each dot represents one participant, horizontal bar represents the group average. 104
- A.4 Trial-level response distributions for Sighted Control and Low Vision groups.** Bars represent the total count of trials for each response value (number of flashes perceived). Left: 1 Flash, 1 Beep (Control Trials) – While Sighted Controls show high detection consistency (peaking at 1), the Low Vision group exhibits a substantial number of “misses” (0 flashes), indicating a lower sensory detection threshold. Right: 1 Flash, 2 Beeps (Illusion Trials) – Inaccuracy in Sighted Controls is driven by a shift toward reporting 2 flashes (the sound-induced flash illusion). In the Low Vision group, inaccuracy is split between “misses” (0) and “illusions” (2). This bimodal error distribution explains why the Low Vision group’s mean perceived flashes can remain near 1.0 despite significantly lower accuracy than controls. 105
- A.5 Illusion strength as a function of stimulus location.** Illusion strength is plotted across eccentricity ($5^\circ - 15^\circ$) and visibility levels (high vs. low). No significant effect of eccentricity or visibility was observed based on mixed-effects modeling, although a slight, non-significant decrease in illusion strength is visible under low-visibility conditions. 105

A.6	Group average Bayesian Causal Inference model fits for all locations. Visual response distributions (solid lines) and corresponding model-predicted distributions (dashed lines) are shown for four conditions (F0B0, F1B0, F1B1, F1B2). The top row depicts group average for low vision participants, and the bottom row depicts group average for sighted controls. Each panel shows the probability of reporting 0–2 flashes for that condition.	106
A.7	Parameter Identifiability Matrix. Correlation matrix for BCI parameters in high-visibility locations. Most parameter pairs showed weak correlations ($ r < 0.4$), including the relationship between p_{common} and σ_V ($r = -0.37$), suggesting that visual uncertainty and binding tendency were largely dissociable within the fitting procedure. A moderate correlation was observed between σ_P and μ_P ($r = 0.54$), indicating partial dependence between prior width and prior mean estimates. Overall, the absence of strong inter-parameter correlations supports reasonable parameter identifiability within the BCI framework, particularly for the key distinction between sensory precision and causal binding tendency.	107
A.8	Parameter recovery analysis. Correlation between ground-truth and recovered parameter values for the BCI model. Synthetic datasets were generated using each participant’s estimated parameters and subsequently re-fitted to evaluate the model’s identifiability and the study design’s robustness. In each panel, the recovered values are plotted against the ground truth (black dashed lines indicate the identity $y = x$). High R^2 and low Mean Squared Error (MSE) were observed for all parameters, indicating these parameters were reliably recovered.	108
A.9	Correlation between Disease Characteristics and BIC Scores (Model Performance; the lower the better). There are no significant correlations between disease duration and BIC score, as well as age of onset and BIC score.	109
A.10	Correlation between Disease Characteristics and Illusion Strength. There are no significant correlations between disease duration and illusion strength (F1B2 - F1B1), as well as age of onset and illusion strength.	109

B.1	Blind Spot Measurements (N = 20). The distribution of distances from the vertical and horizontal meridians, as well as the width and height of the blind spots for each eye (left eye: blue, right eye: peach). Here, the vertical and horizontal meridians refer to the midlines of the monitor along the x- and y-axes. Blind-spot locations and sizes were consistent across individuals. The mean blind-spot area was 21.6deg^2 ($SD = 8.99$). Blind-spot centers were located approximately 15.2° from the fovea ($SD = 0.976$) and 1.69° below the horizontal meridian ($SD = 0.780$).	110
B.2	Eye-tracking measurements during fixation, stimulus presentation, and response phases (N = 20) The y-axis represents the gaze-point distance from the fixation cross (center) in visual degrees, and the x-axis represents time. Solid lines indicate the mean gaze distance across trials, and shaded regions indicate standard deviations. Color code: left eye – blue; right eye – peach. Participants maintained excellent fixation throughout the fixation and stimulus presentation phases. Mean gaze-point distance from the center was 0.016° ($SD = 0.0028$) for the left eye and 0.017° ($SD = 0.0034$) for the right eye during fixation, and 0.017° ($SD = 0.0039$) and 0.018° ($SD = 0.0043$) during stimulus presentation, respectively. Although gaze deviations increased slightly during the response phase (e.g., when reading the text prompt or locating response keys), fixation remained stable during stimulus presentation, confirming that flashes were presented accurately at the intended retinal locations (borders and center of the blind spots). Individual eye-tracking data for all participants are shown in Figure B.3.	111
B.3	Individual eye tracking measurements during fixation, stimulus presentation, and response phases.	112
B.4	Illusion rates in control (F2B2) and illusion (F2B3) conditions. There is a significant increase in 3-flash responses in the illusion condition compared to the control condition. No significant differences are observed between stimulus locations.	113
B.5	Model Comparison by Bayesian Information Criterion (BIC). Mean BIC scores (lower is better) for the four models tested: BCI, MLE, FS, FF. Error bars represent ± 1 standard error of the mean (SEM).	113

- B.6 Parameter recovery analysis.** Correlation between ground-truth and recovered parameter values for the BCI model. Synthetic datasets were generated using each participant's estimated parameters and subsequently re-fitted to evaluate the model's identifiability and the study design's robustness. In each panel, the recovered values are plotted against the ground truth (black dashed lines indicate the identity $y = x$). High R^2 and low Mean Squared Error (MSE) across all panels demonstrate that the model parameters were reliably recovered. 115
- B.7 Parameter Identifiability Matrix.** Pearson correlation coefficients (r) between best-fitting BCI model parameters pooled across participants and locations ($N = 40$). Visual noise (σ_V), numerical prior width (σ_P), and auditory noise (σ_A) show minimal inter-parameter correlation, indicating high identifiability. A significant correlation cluster was observed between the causal prior (P_{common}) and numerical prior mean (μ_P), reflecting a functional coupling between causal belief and numerical expectation in the absence of direct visual input. 116
- C.1 Detailed circuit schematics and PCB layout.** 117
- C.2 Mean accuracy across conditions, all directions pooled.** Error bars represent ± 1 SEM. 117
- C.3 Impact of Spatial Disparity and Visual Reliability on Accuracy.** Mean localization accuracy is shown for the (Left) Visual Target and (Right) Tactile Target tasks. Accuracy is normalized to a $[0, 1]$ scale ($1.0 = 0^\circ$ error); error bars represent ± 1 SEM. A robust three-way interaction was observed ($p < .001$) for visual response, characterized by a dramatic decline in accuracy as spatial disparity increased under low visual reliability (light green). On the other hand, accuracy remained relatively stable and high across visual and spatial manipulations for tactile response. 118
- C.4 Impact of Spatial Disparity, Visual Reliability, and Region on Accuracy.** Mean localization accuracy is shown for the (Left) Visual Target and (Right) Tactile Target tasks. Accuracy is normalized to a $[0, 1]$ scale ($1.0 = 0^\circ$ error); error bars represent ± 1 SEM. We observe similar trends from Fig. C.3. There is no significant 4-way interaction, indicating that these performance trends were consistent across all spatial Regions (Front, Side, and Back). 118

- C.5 **Fitted BCI Parameters.** Parameter values evaluated separately across the Front region under High and Low visual reliability. Large markers represent the group mean (± 1 SEM), with individual participant data overlaid. 119
- C.6 **Response Distributions.** Probability density curves mapping the visual likelihood, tactile likelihood, and spatial prior across the Front region. Consistent with the global fits, the spatial priors remain notably flat across all spatial areas, negating the behavioral relevance of shifts in μ_P 119
- C.7 **Group average MLE model fits.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° 120
- C.8 **Group average FS model fits.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° 121
- C.9 **Group average FF model fits.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° 122

- C.10 Parameter recovery analysis for the Bayesian Causal Inference (BCI) model.** Scatter plots illustrate the correspondence between ground-truth and recovered parameter values across 25 synthetic datasets. To evaluate model identifiability and experimental robustness, synthetic data were generated using each participant's estimated parameters and subsequently re-fitted. In each panel, recovered values are plotted against ground truth, with dashed gray lines representing the identity ($y = x$). Sensory likelihood parameters (σ_V and σ_T) show strong recovery, with σ_V achieving an $R^2 = 0.91$. While p_{common} displays a lower R^2 (0.43), its remarkably low Mean Squared Error (MSE = 0.04) indicates that estimates are highly accurate but clustered within a narrow range, making R^2 sensitive to minor deviations. The relatively poor recovery of spatial prior parameters (σ_P, μ_P)—characterized by low R^2 and high MSE—is expected, as the fitted priors were sufficiently large to be uninformative, rendering the model less sensitive to their specific values. 123
- C.11 Parameter Identifiability Matrices.** Pearson correlation coefficients (r) calculated for the Front region ($N = 25$). The primary parameters ($p_{common}, \sigma_P, \sigma_V$) exhibit minimal inter-parameter correlation, confirming that the BCI framework uniquely identifies these values regardless of spatial location. Observed correlations between visual noise (σ_V) or tactile noise (σ_T) and the spatial prior mean (μ_P) are considered functionally negligible. Given the uninformatively broad spatial prior (σ_P) and the limited recoverability of μ_P , these spatial offsets do not meaningfully influence the resulting perceptual estimates, nor do they compromise the stability of the core causal and sensory parameters. 124

LIST OF TABLES

<i>Number</i>	<i>Page</i>
3.1 Demographic and Clinical Information of Low Vision Cohort	15
3.2 Overview of models and their parameters.	20
3.3 BCI Parameter Estimates and Group Comparisons Across Visibility Levels	29
4.1 Overview of models and their free parameters	45
4.2 BCI Model Parameter Estimates and Statistical Comparisons (N=20)	50
5.1 Overview of models and their free parameters	70
5.2 Performance Metrics for Random Forest Kinematic Classifiers. Classification accuracy and F1 scores are reported for each experi- mental dimension.	76
5.3 Fitted BCI Parameters for Front region. Mean parameter values ± 1 standard error of the mean (SEM) across high and low visual reliability conditions. Angular parameters (σ_V , σ_T , σ_P , μ_P) are expressed in degrees.	81
A.1 BCI Parameter Estimates and Group Comparisons (All Locations) . .	106
B.1 Comparison between three decision strategies of the BCI model (model averaging, probability matching, and model selection). . . .	114

Chapter 1

INTRODUCTION

Sight, hearing, touch, smell, and taste are the five common senses that allow humans to perceive, interpret, and react to the environment. However, as Aristotle first noted in *De Anima*, our experience of the world is not merely a collection of isolated sensations but rather forms a unity [2]. This suggests that human perception is fundamentally multimodal, relying on effective integration (or segregation) of information across sensory modalities. This classical insight is supported by modern findings, demonstrated through phenomena such as the McGurk effect and the sound-induced flash illusion [44, 70], which one sensory modality can significantly influence or even alter the perception of another.

When the senses work in such a tight unity, a critical question arises: what happens when one sense is lost or becomes temporarily unreliable? The brain’s inherent plasticity allows it to adapt through sensory compensation [11, 37]. Functional evidence from blind populations suggests that the brain rewires unused visual architecture to process other modalities [21, 23, 46, 64]. However, behavioral results are mixed; while early blind individuals have enhanced auditory acuity [82] and tactile acuity on fingertips [85], but not on the tongue [17], they also exhibit specific deficits in complex tasks like auditory spatial bisection [32, 80]. This inconsistency suggests that “compensation” is not a simple boost in sensitivity, but a complex re-weighting of how the brain interprets signals. Nevertheless, the sensory substitution (SS) community conducted extensive research on ways of conveying information to the blind about the external environment through other intact senses, such as touch [8, 73] and hearing [45, 83].

This complexity points to a fundamental computational challenge: the Causal Inference Problem. Because the environment is noisy, the brain cannot assume that all concurrent signals belong together. If you hear a bang and see a flash, your brain must decide if they originated from a single event (integration) or two independent sources (segregation). Mathematically, this is often modeled through a Bayesian framework [38], where the observer must choose between two competing hypotheses: (1) $C = 1$, sensory signals come from a common source, (2) $C = 2$, sensory signals come from independent sources. The brain’s decision depends on three

factors: the likelihood (relative reliabilities of each sensory signal), the causal prior (the internal bias that signals originate from a common source), and the perceptual prior (expectations regarding the environment, such as the typical spatial location or frequency of sensory events). The Bayesian framework has been tested empirically across a wide spectrum of perceptual and sensorimotor tasks, such as spatial perception [38, 49, 81, 88], temporal perception [4, 63, 72], spatial-temporal perception [15], weight perception [55], body ownership perception [65], and sensori-motor processing [1, 7, 26, 54].

Under typical conditions, visual signals dominate multisensory processing due to their superior spatial resolution. Indeed, the information channel capacity of the visual system has two orders of magnitude greater bandwidth than audition, and four orders of magnitude greater than touch [34]. This massive disparity in data density explains why the brain typically weighs visual evidence so heavily in perceptual tasks. However, when visual information is removed or degraded, the “computational geometry” of this decision-making process must change. This raises a critical question for the study of neuroplasticity: how does the brain solve the Causal Inference Problem when its most reliable evidence is gone?

The central objective of this dissertation is to isolate the mechanisms underlying this compensation. While it is well-established that the brain exhibits significant plasticity in response to vision loss, it remains unclear whether these changes reflect a fundamental shift in multisensory integration strategy. Does the adaptation to visual uncertainty reflect a modulation in the causal structure (the internal belief in a common cause), or does it demonstrate a flexible re-weighting of sensory likelihoods (relative reliabilities of incoming sensory information)?

This work investigates the limits of sensory compensation across a spectrum of visual uncertainty, spanning chronic impairment, localized absence, and experimentally controlled degradation. To identify the underlying computational mechanisms, we evaluate models of multisensory integration within a Bayesian Causal Inference framework, characterizing how the brain operates as a probabilistic observer under varying input constraints. **Chapter 2** provides an overview of the theoretical frameworks and mathematical formalisms of the computational models tested. **Chapter 3** examines multisensory inference in clinical low vision, using the Sound-Induced Flash Illusion (SIFI) [72] to assess how chronic visual impairment impacts integration across retinal locations. **Chapter 4** investigates the physiological blind spot, evaluating whether the brain integrates signals in a region of complete visual

absence using the Audiovisual Rabbit Illusion [74]. **Chapter 5** extends this work to immersive virtual reality, where visual reliability is systematically manipulated in a visuotactile task to probe how inference strategies adapt in a naturalistic 3D environment. Finally, **Chapter 6** summarizes the key findings and outlines future directions.

Chapter 2

THEORETICAL FRAMEWORK

This chapter formalizes the computational models used to interpret behavioral data in Chapters 3-5. We begin by describing the classical Maximum Likelihood Estimation (MLE) approach, followed by the more flexible Bayesian Causal Inference (BCI), as well as two edge cases for BCI: Forced Fusion (FF) and Full Segregation(FS), where observers always integrate or segregate, respectively.

Maximum Likelihood Estimation (MLE)

This model was first proposed by Ernst and Banks [29]. They posited that humans integrate sensory signals in a statistically optimal fashion – in order to minimize variance in the final estimate, the more reliable sensory modality dominates the final percept. The MLE model serves as the optimal baseline for sensory integration. A key assumption of this model is that the brain always treats concurrent signals as originating from a common source.

Given a visual stimulus s_V and a tactile stimulus s_T , the brain’s internal representations (x_V, x_T) are modeled as Gaussian distributions centered at the true stimulus location, with variances σ_V^2 and σ_T^2 .

$$x_V \sim N(s_V, \sigma_V), \quad x_T \sim N(s_T, \sigma_T) \quad (2.1)$$

The integrated estimate $\hat{S}_{VT}$ is obtained under the assumption of a squared loss function by combining the visual and tactile estimates (x_V, x_T) weighted by their relative reliabilities (σ_V, σ_T) :

$$\hat{S}_{VT} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_T}{\sigma_T^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_T^2}} \quad (2.2)$$

The primary prediction of MLE is that the variance of the combined estimate (σ_{VT}^2) will be lower than the variance of either single modality:

$$\sigma_{VT}^2 = \frac{\sigma_V^2 \sigma_T^2}{\sigma_V^2 + \sigma_T^2} \quad (2.3)$$

To illustrate the “shrinkage” of the variance of the combined estimate, two scenarios were simulated: (1) When visual and tactile modalities are equally reliable i.e., same

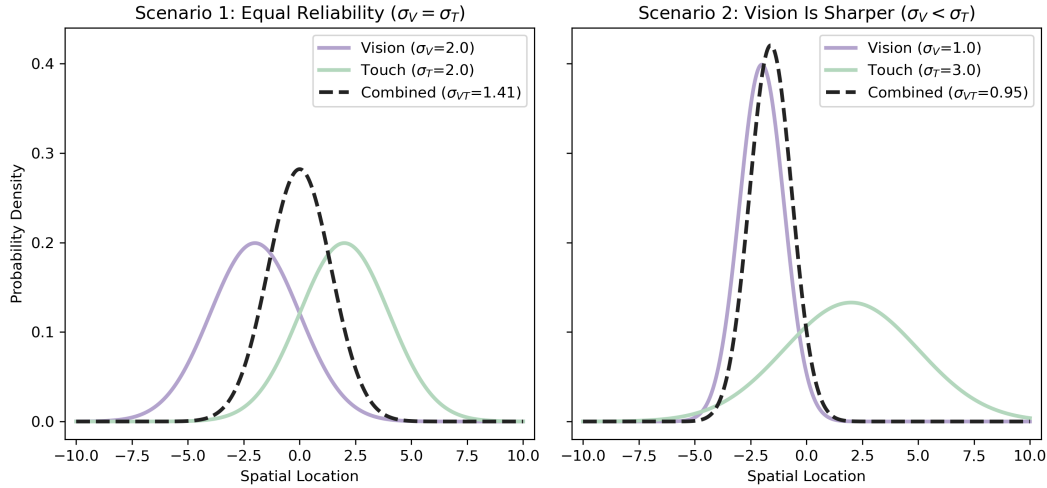


Figure 2.1: **Visualization of MLE Model.** Left: Model behavior when $\sigma_V = \sigma_T$. Right: Model behavior when $\sigma_V < \sigma_T$.

σ_s , (2) when visual is more reliable i.e., $\sigma_V < \sigma_T$ (Fig. 2.1). Because the variance of the combined estimate is lower than the variance of either single modality, as an “ideal observer”, the optimal strategy is to always combine information across sensory modalities to minimize variance, hence the model assumption that humans always integrate.

Bayesian Causal Inference (BCI)

When sensory cues are in proximity to each other in time, space, and structure, there is strong evidence for cues sharing a common cause. However, when cues show a clear disparity, it is no longer ecologically optimal for the brain to integrate them. In these cases, MLE cannot account for behavior because it lacks a mechanism for segregation. To address this, Körding et al. [38] proposed a generative model – Bayesian Causal Inference (BCI) – that treats the causal structure of the environment as a latent variable C to be inferred.

The BCI framework assumes that the observer must choose between two competing hypotheses: (1) Common Cause ($C = 1$): Both visual (s_V) and tactile (s_T) stimuli originate from a single source (s). In this case, the signals are integrated according to the MLE rule; (2) Independent Causes ($C = 2$): Each stimulus originates from a distinct source ($s_V \neq s_T$). In this case, the signals are processed independently, leading to full segregation (Fig. 2.2). The observer’s internal representations (x_V, x_T) are corrupted by Gaussian noise, just as in the MLE model (Equation 2.1).

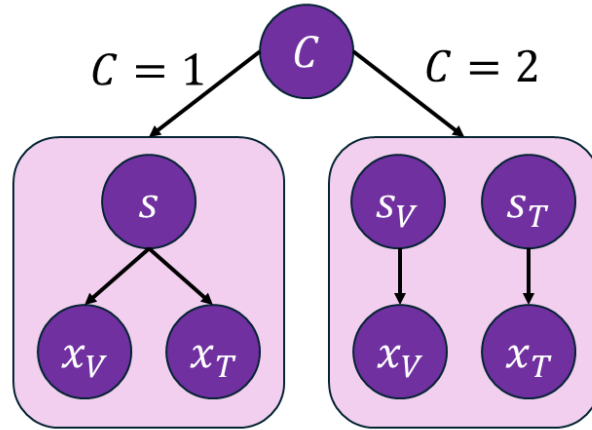


Figure 2.2: **Visualization of BCI Model.** Left: Model behavior when the common cause ($C = 1$) hypothesis is assumed. Right: Model behavior when the independent cause ($C = 2$) hypothesis is assumed.

However, the observer now calculates the posterior probability of a common cause by combining the likelihood of the observed data with an internal causal prior, (P_{common} ; tendency to integrate signals):

$$P(C = 1|x_V, x_T) = \frac{p(x_V, x_T|C = 1)P_{common}}{p(x_V, x_T|C = 1)P_{common} + p(x_V, x_T|C = 2)(1 - P_{common})} \quad (2.4)$$

In addition to the causal prior P_{common} , the BCI model also includes a spatial prior that represents the prior spatial bias for where stimuli arise. Similar to sensory likelihoods (Equation 2.5), it is corrupted by Gaussian noise.

$$x_P \sim N(\mu_P, \sigma_P) \quad (2.5)$$

The calculation for optimal estimates is similar to that of MLE (Equation 2.2), with the inclusion of the spatial prior term. In the case of a common cause ($C = 1$), the optimal visual-tactile estimate $\hat{s}_{(VT,C=1)}$ is given in Equation 2.6. In the case of an independent cause ($C = 2$), the optimal estimate of the unisensory visual $\hat{s}_{(V,C=2)}$ and tactile $\hat{s}_{(T,C=2)}$ stimuli are independent and given in Equation 2.7.

$$\hat{s}_{(VT,C=1)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_T}{\sigma_T^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_P^2}} \quad (2.6)$$

$$\hat{s}_{(V,C=2)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}, \hat{s}_{(T,C=2)} = \frac{\frac{x_T}{\sigma_T^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_T^2} + \frac{1}{\sigma_P^2}} \quad (2.7)$$

Because the brain often faces uncertainty regarding the causal structure (i.e., $P(C = 1|x_V, x_T) \approx 0.5$), it must employ a decision strategy to combine the estimates under the two causal structures and produce a final estimate. At present, there are three widely used decision strategies for BCI modeling [88] known as “model averaging”, “model selection”, and “probability matching”.

If the objective is to minimize the mean squared error (MSE) of the spatial estimates, the optimal strategy is *model averaging*. In this approach, the brain does not commit to a single causal structure but instead weights the estimates from both hypotheses ($C = 1$ and $C = 2$) according to their relative posterior probabilities. The final visual $\hat{s}_V$ and tactile $\hat{s}_T$ estimates are calculated as:

$$\begin{aligned} \hat{s}_V &= p(C = 1|x_V, x_T)\hat{s}_{(VT,C=1)} + p(C = 2|x_V, x_T)\hat{s}_{(V,C=2)} \\ \hat{s}_T &= p(C = 1|x_V, x_T)\hat{s}_{(VT,C=1)} + p(C = 2|x_V, x_T)\hat{s}_{(T,C=2)} \end{aligned} \quad (2.8)$$

This strategy leads to “partial integration”, where the influence of the second modality decreases gracefully as the spatial disparity increases.

In contrast, *model selection* assumes the brain chooses the single most likely causal structure and ignores the other. If $P(C = 1|x_V, x_T) > 0.5$, the observer reports the fused estimate $\hat{s}_{VT,C=1}$. If the probability is lower, the observer reports the unisensory estimate i.e., $\hat{s}_{V,C=2}$. This strategy maximizes consistency between the causal structure chosen and the estimate of location.

$$\begin{aligned} \hat{s}_V &= \begin{cases} \hat{s}_{VT,C=1}, & \text{if } p(C = 1|x_V, x_T) > 0.5 \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_T &= \begin{cases} \hat{s}_{VT,C=1}, & \text{if } p(C = 1|x_V, x_T) > 0.5 \\ \hat{s}_{T,C=2}, & \text{otherwise} \end{cases} \end{aligned} \quad (2.9)$$

Finally, *probability matching* suggests that the observer chooses a causal structure proportionally to its posterior probability. For example, if there is a 70% chance of a common cause, the observer will use the integrated estimate 70% of the time and the segregated estimate 30% of the time. This strategy is considered to be sub-optimal in terms of economic and utility theory because once it is known that

the visual-tactile stimuli with common causes are presented more often, observers should report the fused estimate on all trials to maximize their utility or gain (.70 proportion correct vs $.70^2 + .30^2 = .58$ proportion correct).

$$\begin{aligned}\hat{s}_V &= \begin{cases} \hat{s}_{V,C=1}, & \text{if } p(C = 1|x_V, x_T) > \xi, \xi \sim U(0, 1) \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_T &= \begin{cases} \hat{s}_{T,C=1}, & \text{if } p(C = 1|x_V, x_T) > \xi, \xi \sim U(0, 1) \\ \hat{s}_{T,C=2}, & \text{otherwise} \end{cases}\end{aligned}\tag{2.10}$$

Boundary Cases: Forced Fusion (FF) and Full Segregation (FS)

To validate the necessity of the BCI framework, we test it against two restricted boundary cases. These models are mathematically nested within the BCI framework by “clamping” the P_{common} parameter.

Forced Fusion (FF) occurs when P_{common} is fixed to 1.0. Under this constraint, the brain ignores the competing causal structure (i.e., segregation) and forces integration regardless of how far apart the visual and tactile signals are in space or time. The posterior probability $P(C = 1|x_V, x_T)$ (Equation 2.4) becomes a constant 1, and the observer’s behavior is dictated entirely by the MLE weighting rule (Equation 2.6). Forced Fusion is functionally equivalent to MLE i.e. always integrates, but the key difference is that the integrated estimate is not just a weighted average of unisensory estimates, but includes a prior term $p(s)$. It is important to note that if σ_P is very large (a “flat” prior), the prior becomes uninformative, thus Equation 2.6 simplifies back to the standard MLE estimate (Equation 2.2).

The *Full Segregation (FS)* model represents the opposite boundary case, where the causal prior is fixed to $P_{common} = 0.0$. In this scenario, the brain treats the visual and tactile signals as originating from entirely independent sources ($C = 2$). Consequently, the posterior probability of a common cause (Equation 2.4) is always 0. Under Full Segregation, no multisensory integration occurs. However, similar to the FF model, the FS model incorporates the prior term $p(s)$. This means that even though the visual and tactile signals do not influence each other, they are both still pulled toward the observer’s internal expectation of stimulus location (Equation 2.7).

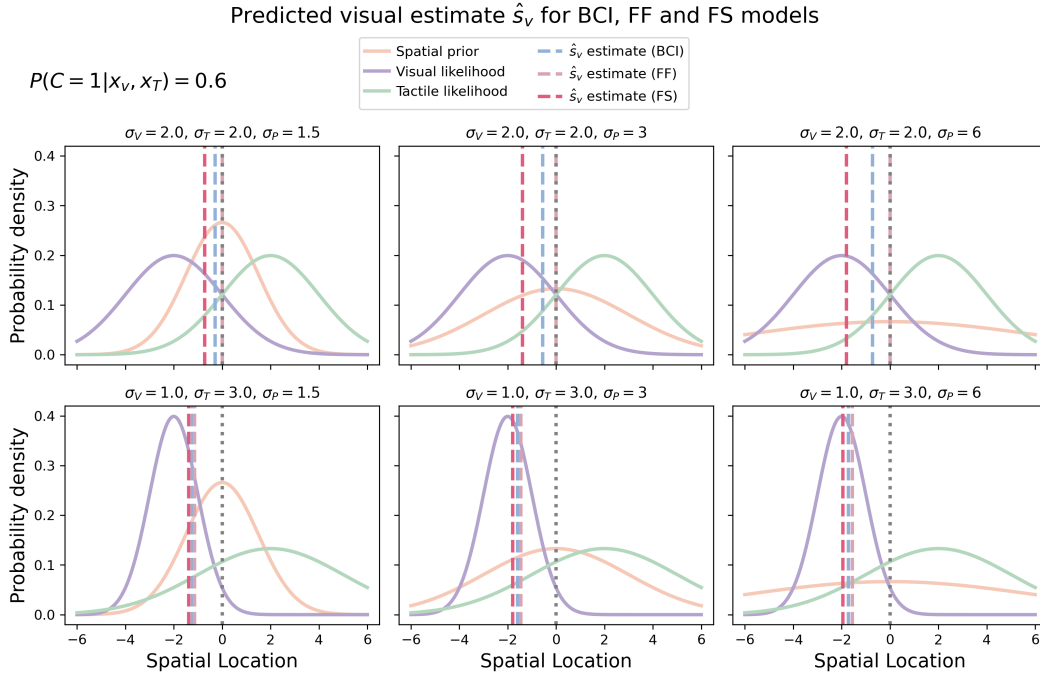


Figure 2.3: **Simulated visual estimates under varying sensory reliability and spatial priors.** Each panel displays the spatial prior ($p(s)$, peach), the visual likelihood (x_V , purple), and the tactile likelihood (x_T , green). The dashed vertical line indicates the predicted visual estimate ($\hat{s}_V$) under the BCI model using model averaging, FF, and FS models. Posterior probability of a common cause $P(C = 1 | x_V, x_T)$ is fixed at 0.6. **Top Row:** Equal reliability ($\sigma_V = \sigma_T = 2.0$). **Bottom Row:** Tactile uncertainty exceeds visual uncertainty ($\sigma_V = 1.0, \sigma_T = 3.0$), simulating conditions of high visual reliability. **Columns (Left to Right):** The spatial prior becomes progressively flatter ($\sigma_P = 1.5, 3, 6$), demonstrating a decreasing “central bias” and increasing reliance on sensory data.

To illustrate the distinct computational architectures of the Bayesian Causal Inference (BCI), Forced Fusion (FF), and Full Segregation (FS) models, Figure 2.3 presents simulated probability density functions and visual estimates $\hat{s}_V$ for different parameter values. Each panel displays the visual likelihood (x_V , purple), the tactile likelihood (x_T , green), and the spatial prior ($p(s)$, peach). Posterior probability of a common cause $P(C = 1 | x_V, x_T)$ is fixed at 0.6. The top row simulates conditions of equal sensory reliability ($\sigma_V = \sigma_T = 2.0$), while the bottom row simulates conditions of high visual reliability ($\sigma_V = 1.0, \sigma_T = 3.0$). As the spatial prior becomes flatter across the columns (from $\sigma_P = 1.5$ to 6), its “pull” toward the center weakens.

The dashed lines indicate the predicted final visual estimate ($\hat{s}_V$) for each model. The FS model (magenta line) serves as the non-integrated baseline, predicting a weighted percept strictly between the visual likelihood and the peach prior (Equation 2.7). The FF model (rose line) represents absolute integration, always predicting a weighted

percept between the visual, tactile, and spatial prior (Equation 2.6). The BCI model (using model averaging, blue line) provides the most flexible prediction. Under this strategy (Equation 2.8), the BCI percept is not a simple MLE average but a weighted compromise between the integrated (FF) and segregated (FS) solutions, weighted by the posterior probability of a common cause (in this case, $P(C = 1|x_V, x_T) = 0.6$). By comparing these three lines, we can isolate distinct components of a participant’s behavioral bias. The distance between the true visual source and the FS estimate (magenta line) quantifies the “centrality bias” caused by the spatial prior. Meanwhile, the further shift seen in the BCI and FF estimates (blue and rose lines) quantifies the “multisensory bias” caused specifically by the tactile cue as well as prior belief in a common cause in BCI. This computational decomposition allows us to determine whether a participant’s report is being “pulled” toward the center by their internal expectations or toward the tactile cue by a belief in a common source.

While Forced Fusion (FF) and Full Segregation (FS) represent the theoretical limits of multisensory behavior, they serve as critical statistical baselines for evaluating the Bayesian Causal Inference (BCI) framework. In experimental contexts, FF and FS act as null models: FF assumes a system that maximizes precision at the cost of spatial accuracy during disparity, whereas FS assumes a system that prioritizes signal independence. By treating these two models as nested variants of the BCI framework (where the causal prior P_{common} is constrained to 1 or 0, respectively), we can employ the Bayesian Information Criterion (BIC) to adjudicate between them. This approach ensures that the added complexity of the BCI model (i.e., the free parameter P_{common}) is only accepted if it provides a significantly better fit to the participant’s data than the simpler, constrained alternatives. Consequently, this comparison allows us to quantitatively determine whether the human brain adopts a fixed integration strategy or, as the BCI model suggests, a flexible strategy that accounts for the underlying causal structure of the environment.

Chapter 3

MULTISENSORY INFERENCE IN LOW VISION

The following chapter’s contents are taken and adapted from Chan et al. 2025, with modifications done to fit the dissertation format.

Ailene Y. C. Chan, Noelle R. B. Stiles, Carmel A. Levitan, Armand R. Tanguay Jr., Shinsuke Shimojo (2025). Did You See the Sound? A Bayesian Perspective on Crossmodal Perception in Low Vision. In: *bioRxiv*
DOI: 10.64898/2025.12.24.696433.

Abstract

Multisensory integration is often assumed to increase when visual input is degraded, yet it remains unclear whether low vision enhances susceptibility to cross-modal illusions or whether such effects depend on local variations in visual reliability. We tested low vision and sighted control participants on the Sound-Induced Flash Illusion (SIFI) across 24 visual-field locations, concurrently measuring baseline flash-detection accuracy. While both groups showed the expected auditory-driven increase in perceived flash numerosity, sighted controls were significantly more susceptible to the classic “double-flash” percept. While illusion strength approached marginal significance for eccentricity, it was more robustly predicted by local flash-detection accuracy, indicating that sound-induced percepts depend on the availability of a reliable visual signal. Exploratory Bayesian Causal Inference (BCI) modeling revealed that when fitted across all locations, model performance was systematically weaker and more variable for low vision observers. Crucially, however, restricting the analysis to high-visibility locations yielded robust model fits ($R^2 \geq 0.97$). This suggests that, like sighted controls, low vision participants continue to behave in a manner broadly consistent with Bayesian-optimal inference when visual information is sufficiently reliable. Together, these findings demonstrate that low vision does not globally amplify audiovisual interactions; rather, multisensory enhancements are constrained by local visual reliability, and the “Bayesian brain” remains computationally intact only when provided with a sufficient threshold of sensory input.

3.1 Introduction

In every waking moment, the nervous system must translate incomplete and noisy sensory input into a coherent percept. When multiple senses are simultaneously perceived the brain must also infer whether multisensory information originates from a single cause or from separate events. A healthy nervous system accomplishes this by weighing sensory cues according to their reliability and by incorporating prior beliefs (or heuristics) about how sensory events unfold [29, 38]. In this framework, perception reflects the statistically optimal combination of cues given their uncertainties and the observer’s expectations.

But how do these computational principles adapt to chronic sensory degradation? While multisensory integration is well-characterized in healthy observers, much less is known about how these mechanisms operate when a primary sense is pathologically compromised. Previous research using experimental manipulations of visual contrast in healthy sighted participants has shown that while the visual likelihood (σ_V) adjusts to accommodate reduced reliability, internal priors (e.g., the prior for common causality, p_{common}) remain remarkably stable [10]. This stability is perhaps expected in healthy observers whose daily experience is dominated by high-fidelity visual input.

However, it remains unclear whether these principles hold when visual degradation is not an experimental manipulation, but rather a permanent, pathological state. Low vision offers a particularly informative case: rather than a complete loss of vision, patients retain islands of intact visual function interspersed with regions of reduced sensitivity. This raises an important question: does the visually impaired brain globally reorganize its multisensory strategy, or does it continue to adjust integration locally based on the immediate reliability of the visual signal at specific retinal locations?

Neural evidence suggests that the visually impaired brain undergoes cross-modal reorganization, with auditory and tactile signals recruiting occipital cortex even when some visual function remains [23, 46]. Anatomical studies further reveal direct projections from auditory cortex to early visual regions [9, 30, 57], especially representing peripheral visual space. These findings raise the possibility that cross-modal influences on perception may be enhanced in low vision. However, behavioral findings regarding this enhancement remain mixed. Some evidence suggests that sound can facilitate visual detection in the low vision group [76], yet other studies have found that individuals with severe visual impairment perform worse on tasks

requiring complex multisensory integration [47]. It remains a mystery whether such possible enhancement (or reduction) depends on global differences in sensory processing (e.g., reduced weighting of vision) or local differences driven by the visibility of specific retinal regions.

To address this gap, we used the Sound-Induced Flash Illusion (SIFI) [70], a classic demonstration in which auditory beeps alter the perceived number of visual flashes. We presented the illusion at 24 locations across the visual field in both low vision and sighted control participants. Our study aimed to determine: (i) if low vision participants exhibit altered susceptibility to the illusion compared to controls; (ii) how multisensory interactions vary across the visual field; and (iii) whether illusion strength is tied to local visual sensitivity.

While the BCI framework has been successfully applied to healthy multisensory perception [4, 63, 69], its applicability to chronic clinical populations remains less explored. We therefore adopted an exploratory computational approach, applying BCI modeling to different spatial partitions of the visual field. This allowed us to formally test whether the increased illusion susceptibility often attributed to low vision is a product of altered prior expectations (p_{common}) or a direct consequence of increased sensory uncertainty (σ_V) within a preserved Bayesian architecture.

3.2 Methods

Human Participants

We first recruited 21 neurotypical participants (11 females; mean age = 26.3 years, range = 18–40 years) from the California Institute of Technology community. All had normal or corrected-to-normal vision (20/20) and hearing. Recruitment took place during the early phase of the study, when COVID-19 restrictions limited access to external participants.

We later recruited 25 individuals with low vision after obtaining written informed consent, either through an e-reader or verbally through the experimenter. Four participants did not complete the full experimental protocol and were excluded from data analysis. The final low-vision cohort included 21 participants (8 females; mean age = 51.5 years, range = 25–81 years), all with normal or corrected-to-normal hearing. Demographic and clinical details are shown in Table 3.1. To account for variations in visual field sensitivity across participants, particularly given the lack of recent clinical perimetry for low vision individuals, we implemented a custom Visual Flash Detection Task. This task evaluated detection accuracy at each of the

24 stimulus locations used in the SIFI Task. These data (provided in Supplementary Data Set 1) served as a functional baseline to verify visibility at each stimulus location as well as the eccentricity levels tested (5°, 10°, and 15°).

After testing the low-vision group, we recruited an additional cohort of age-matched neurotypical controls to better match the age distribution of the low-vision participants. Five participants from the original control cohort were retained because their ages already fell within the target range of the low-vision group and therefore met the age-matching criteria. To complete the age-matched sample, 16 new control participants were recruited, resulting in a final control cohort of 21 participants (15 females; mean age = 51.0 years, range = 24–80 years).

Participants were compensated at \$25/hour. All procedures were approved by the institutional review board at the California Institute of Technology (IRB IR20-1025).

Simuli and conditions

Visual Flash Detection Task

Visual stimuli were circular white discs (2° in diameter) presented on a black background at 100% contrast for one frame (16.7 ms at 60 Hz). Stimuli appeared at 24 possible locations, corresponding to eight equally spaced positions along the circumference at 5°, 10°, and 15° eccentricities from central fixation (Fig. 3.1A). Two audiovisual conditions were tested in this task, following the naming convention FXBY, where X denotes the number of flashes and Y denotes the number of beeps such as F0B0 (no flashes or beeps) and F1B0 (one flash, no beeps). The stimulus timing for these conditions is illustrated in Fig. 3.1B–C.

SIFI Task

This task used the same visual stimuli and stimulus locations as described above (Fig. 3.1A). Auditory stimuli were pure tones (800 Hz, 7 ms duration). Two audiovisual conditions were tested: F1B1 and F1B2. The F1B2 condition, also known as the Sound-Induced Flash Illusion condition, elicits the perception of a second illusory flash induced by the second auditory beep. Timing of the stimuli is shown in Fig. 3.1D–E.

Table 3.1: Demographic and Clinical Information of Low Vision Cohort

Participant	Diagnosis	Age (Years)	Sex	Age of Onset (Years)	Duration (Years)
LV001	Retinitis pigmentosa	65	F	31	34
LV002	Microphthalmia	43	M	0	43
LV003	NAION, right eye	74	M	68	6
LV005	Congenital cataracts	69	M	0	43
LV007	Macular degeneration	70	F	68	2
LV008	Choroideremia	33	M	23	10
LV009	Stargardt's macular de- generation	28	M	14	14
LV011	Left Phthisis Bulbi, Right cystoid macular edema, uveitis	66	F	24	42
LV012	Glaucoma, right eye to- tal blind	56	M	42	14
LV013	Glaucoma, right eye to- tal blind	40	M	35	5
LV015	Retinitis pigmentosa, detached retina, right eye only	57	M	11	46
LV016	Chronic Relapsing In- flammatory Optic Neu- ritis (CRION)	52	F	36	16
LV017	Enlarged optic nerves	33	M	27	6
LV018	Dominant optic atrophy	25	F	12	13
LV019	Glaucoma, left eye only	81	M	76	5
LV020	Retinal degeneration	42	M	0	42
LV021	Retinitis Pigmentosa	41	F	21	20
LV022	NAION	64	M	56	8
LV023	Undeveloped optic nerve	57	F	0	57
LV024	Optic nerve damage due to trauma	42	F	3	39
LV025	Microphthalmia	44	M	0	44

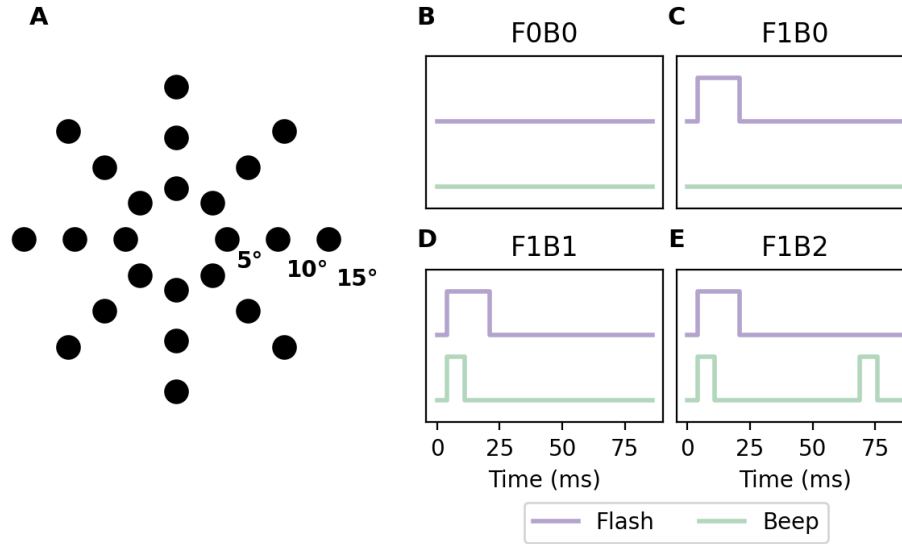


Figure 3.1: **Stimuli and conditions** (A) Stimulus locations corresponding to eight equally spaced positions along the circumference at 5°, 10°, and 15° eccentricities from central fixation. (B-E) Schematic of the four main experimental conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps). When paired, visual flashes (16.7 ms) and auditory beeps (7 ms) had simultaneous onsets.

Experimental Design

Visual Flash Detection Task

The task was performed monocularly for each eye in randomized order, with the non-tested eye covered by an eyepatch. Participants indicated whether they detected a flash (*Yes*: press left arrow button; *No*: press right arrow button). The F1B0 condition was presented five times per stimulus location, and the F0B0 condition served as a single catch trial. All trials were randomized.

SIFI Task

The task was also performed separately for each eye with a randomized order, with the non-tested eye covered by an eyepatch. Participants reported the number of perceived flashes (0–3). The prompt “How many flash(es)?” was designed to be linguistically neutral, explicitly allowing for both singular and plural percepts without biasing the participant toward either. The F1B2 condition was presented ten times per stimulus location, and the F1B1 condition served as a single catch trial. All trials were randomized.

Experimental setup

Stimuli were presented using Psychtoolbox 3.0.17 in MATLAB 2021a. Auditory tones (62 dB SPL) were delivered via stereo speakers (Bose Companion 2 Series III) placed beside the monitor. Visual stimuli appeared on a 27-inch LCD display (Dell UltraSharp U2720Q; 60 Hz). Participants sat 50 cm from the display with their head stabilized on a chin rest and the non-tested eye occluded. The experiment took place in a darkened room. All experimental scripts are publicly available under a CC-BY license [14].

Data analysis

Average Perceived Flash Count by Task and Stimulus Location

To visualize overall perceptual patterns, we first computed the mean number of flashes reported at each stimulus location for both the Visual Flash Detection (F1B0 trials) and SIFI (F1B2 trials) tasks (Fig. 3.2). Locations were further categorized individually for each participant as high-visibility (flash detection accuracy > 50%) or low-visibility ($\leq 50\%$) based on performance in the detection task (i.e., correctly reporting the presence and absence of a flash in F1B0 unimodal trials). To establish the baseline visual accuracy and the presence of the illusion, two Wilcoxon signed-rank tests were conducted within each group to compare perceived flash counts between the baseline (F1B0) and the illusion (F1B2) conditions, with data pooled across all stimulus locations and both eyes. To control for multiple comparisons, Bonferroni correction was applied across the two tests.

Accuracy

Accuracy was defined as the proportion of trials in which participants correctly reported the true number of flashes presented (Fig. 3.3A). To examine how visual and spatial factors influenced visual accuracy, we fit two separate linear mixed-effects models using the *statsmodels* package in Python. In both models, the dependent variable was mean flash perception accuracy and we included a random intercept for participants to account for individual differences in overall accuracy. The first model tested the effects of *group* (Low Vision, Sighted Control), *condition* (F0B0, F1B0, F1B1, F1B2), and *visibility* (High-visibility, Low-visibility), along with all interactions among these factors. The second model tested the effects of *group*, *condition*, and spatial *eccentricity* (5°, 10°, 15°), and their interactions.

Flash Detection vs. Perceived Flash Count (F1B1, F1B2)

We examined whether sounds influenced perceived flash numerosity correlates with visual sensitivity (i.e., flash detection accuracy in the F1B0 condition). For each participant, eye, and stimulus location, we first computed flash-detection accuracy in the F1B0 condition, yielding 48 data points per participant (24 locations $\times$ 2 eyes). We then computed, separately for each beep numerosity (B1, B2), the mean number of flashes perceived at each location and eye, again yielding 48 data points per beep condition. To assess the relation between visual sensitivity and sound-driven changes in perceived flash numerosity, we correlated flash-detection accuracy with the mean number of flashes reported using Pearson’s correlation coefficient, computed separately for each group (Low-Vision, Sighted Control) and beep condition (B1, B2).

Model Fitting

Participants’ responses were modeled using a Bayesian Causal Inference (BCI) framework [38], from which the forced-fusion, full-segregation, and maximum-likelihood estimation (MLE) models [29] were derived as special cases. Model fitting was implemented via the BCI Toolbox [89] using its API for automated batch processing. To ensure parameter convergence and resolve discrepancies between the distributed package and the source code, we implemented several custom bug fixes; our modified version of the toolbox and all associated analysis scripts are publicly available under a CC-BY licensee [14].

In the BCI model, each sensory event (s_i) gives rise to a noisy observation (x_i), sampled from a Gaussian centered on the true stimulus value with modality-specific uncertainty (σ_V , σ_A ; Eq. 3.1). Observers also hold a prior expectation about stimulus numerosity, modeled as a Gaussian with mean μ_P and standard deviation σ_P (Eq. 3.2).

$$x_V \sim N(s_V, \sigma_V), \quad x_A \sim N(s_A, \sigma_A) \quad (3.1)$$

$$x_P \sim N(\mu_P, \sigma_P) \quad (3.2)$$

Because these Gaussian distributions are continuous, we mapped the model’s internal estimates to the discrete nature of the experimental task. Following the BCI toolbox implementation, internal estimates were rounded to the nearest integer and clipped to the range of possible response values [0, 3].

The model infers whether visual and auditory signals arise from a common ($C = 1$) or independent ($C = 2$) cause by combining sensory likelihoods with a causal prior (p_{common}) via Bayes' rule (Eq. 3.3). When signals are inferred to share a common cause, the perceived numerosity is a reliability-weighted average of the visual, auditory, and prior estimates; if inferred as independent, each modality is estimated separately.

$$p(C = 1|x_V, x_A) = \frac{p(x_V, x_A|C = 1)p_{\text{common}}}{p(x_V, x_A|C = 1)p_{\text{common}} + p(x_V, x_A|C = 2)(1 - p_{\text{common}})} \quad (3.3)$$

Observers combine these estimates across causal structures ($C = 1, 2$) using one of three decision strategies: *model averaging* (weighted by posterior probability), *model selection* (choosing the more probable cause), or *probability matching* (stochastic selection in proportion to posterior probability) [88].

Parameter Constraints and Validation

To maintain model parsimony, several parameters were fixed based on empirical evidence. The auditory likelihood (σ_A) was set to 0.2, consistent with recent work using identical stimulus configurations in spatio-temporal variants of the sound-induced flash illusion [15, 74]. To validate this choice for our specific population, we measured beep-detection thresholds in a subset of age-matched controls ($N = 16$). The resulting mean fitted likelihood ($\sigma_A = 0.19 \pm 0.08$) was highly consistent with our fixed value (see Appendix S1). Because only flash responses were collected from the main experimental group, auditory responses were simulated by drawing from a Gaussian distribution centered at the true stimulus count (Eq. 3.1) using the fixed σ_A .

To investigate the spatial stability of these processes, we fitted each of the four models (BCI, forced-fusion, full-segregation, and MLE) across six distinct stimulus location subsets for each participant: (i) all locations combined, (ii) high-visibility locations, (iii) low-visibility locations, and (iv) eccentricities at 5° , 10° , and 15° . This hierarchical approach allowed us to evaluate model performance as a function of both sensory reliability and eccentricity. Free and fixed parameters for each model variant are summarized in Table 3.2.

Table 3.2: Overview of models and their parameters.

Model	Free Parameters	#	Fixed Parameters
Bayesian Causal Inference	$p_{\text{common}}, \sigma_P, \mu_P, \sigma_V$	4	σ_A
Forced-fusion	$\sigma_P, \mu_P, \sigma_V$	3	$\sigma_A, p_{\text{common}} = 1$
Full-segregation	$\sigma_P, \mu_P, \sigma_V$	3	$\sigma_A, p_{\text{common}} = 0$
Maximum-likelihood estimation	σ_V	1	$\sigma_A, p_{\text{common}} = 1, \sigma_P, \mu_P$

Note: Fixed parameters were set to $\sigma_A = 0.2$ for all models. For MLE, $\mu_P = 1.5$ and $\sigma_P = 4000$.

Model Evaluation

Model performance was evaluated using the Bayesian Information Criterion (BIC) for model selection and the coefficient of determination (R^2) for goodness-of-fit. BIC was prioritized to account for varying model complexity ($BIC = k \ln(n) - 2 \ln(\hat{L})$), while R^2 quantified the proportion of explained variance across all experimental conditions:

$$R^2 = 1 - \frac{\sum (y_{\text{obs}} - y_{\text{pred}})^2}{\sum (y_{\text{obs}} - \bar{y}_{\text{obs}})^2} \quad (3.4)$$

where y_{obs} and y_{pred} represent the observed and predicted response proportions across all experimental conditions, respectively.

The robustness of the fixed auditory parameter was verified through a parameter sweep ($\sigma_A \in \{0.1, 0.2, 0.3, 0.4, 0.6, 0.8\}$; see Supplementary Fig. A.2). Group-level fits peaked within the 0.1 – 0.3 range, with a sharp decline in goodness-of-fit for values ≥ 0.4 . Statistical comparisons (Wilcoxon signed-rank tests, Bonferroni-corrected) confirmed that $\text{Model}_{\sigma_A=0.3}$ performed significantly worse than $\text{Model}_{\sigma_A=0.2}$ across both BIC and R^2 metrics ($p_{\text{adj}} < .001$). While $\text{Model}_{\sigma_A=0.1}$ yielded lower BIC scores than $\text{Model}_{\sigma_A=0.2}$ ($p_{\text{adj}} < .001$) for the control group, it offered no significant improvement in R^2 ($p_{\text{adj}} = 0.26$). For the low vision group, $\text{Model}_{\sigma_A=0.1}$ offered no improvement (both BIC and R^2 $p_{\text{adj}} = 1$). Given this lack of increased predictive power and the alignment of $\text{Model}_{\sigma_A=0.2}$ with our control group’s empirical thresholds, we retained $\text{Model}_{\sigma_A=0.2}$ for all primary analyses.

Finally, we assessed whether stimulus visibility or spatial location constrained model interpretability. To allow for direct comparison across datasets of varying sizes, normalized BIC scores (BIC/n) were calculated for every model variant at each spatial location group. For each participant and location subset, this was defined as the raw BIC score divided by the total number of trials (n) included in that specific fit (e.g., $BIC/300$ for a 300-trial subset).

3.3 Results

Average Perceived Flash Count

Participants perceived more flashes during the Illusory Double Flash (F1B2) task (low vision: 1.03 ± 0.15 ; sighted control: 1.61 ± 0.1) than during the Visual Flash Detection (F1B0) task (low vision: 0.55 ± 0.07 ; sighted control: 0.98 ± 0.01), consistent with an auditory-induced enhancement of visual perception (Fig. 3.2, A.3). This effect was significant in both groups (both Bonferroni-corrected $p < .001$). However, only sighted control participants exhibited the full Sound-Induced Flash Illusion (SIFI), reporting an average of 1.61 ± 0.1 flashes across all stimulus locations. Low vision participants reported an average of 1.03 ± 0.15 flashes.

Together, these findings indicate that while auditory cues increased perceived flash counts in both groups, the canonical SIFI—perceiving two flashes in response to one flash paired with two beeps—was reliably experienced only by sighted controls.

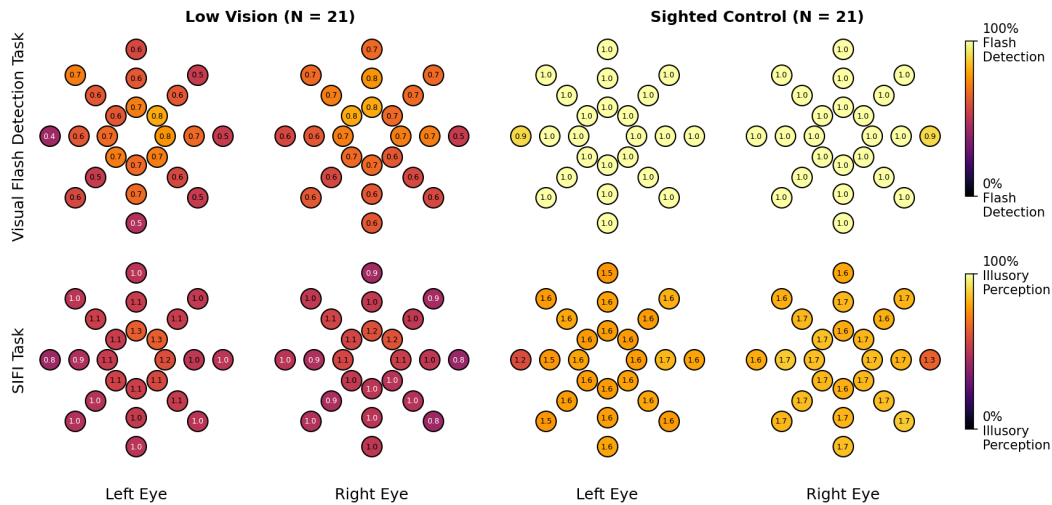


Figure 3.2: **Mean number of flashes perceived by task across all stimulus locations.** Top row: *Visual Flash Detection task* (F1B0 condition). Bottom row: *Illusory Double Flash task* (F1B2 condition). Left panel: Low-vision group. Right panel: Sighted control group. Stimulus locations were further categorized for each participant as **high-visibility** (flash detection accuracy $> 50\%$) or **low-visibility** ($\leq 50\%$) based on performance in the detection task.

Accuracy

We defined accuracy as the proportion of trials in which participants correctly reported the exact number of *flashes* presented (0 or 1; Fig. 3.3A). For F1BY conditions (one physical flash present), it is important to note that accuracy reflects a combination of “missed” responses (0 flashes perceived), “correct” responses (1

flash perceived), and illusory responses (2 or 3 flashes perceived). To examine how auditory context, spatial location, and group membership influenced these measures, we fit two linear mixed-effects models with *participant identity* included as a random effect. Both models used *accuracy* as the dependent variable. Fixed effects included *group* (low vision, sighted control), *condition* (F0B0, F1B0, F1B1, F1B2), and *stimulus location*.

The spatial fixed effects were structured as follows:

- **Model One:** Locations were binned by *functional visibility* (high-visibility vs. low-visibility), defined by each participant's baseline performance (Fig. 3.3B).
- **Model Two:** Locations were binned by *eccentricity* (5° , 10° , 15°), to assess the effect of retinal position regardless of individual visibility thresholds (Fig. 3.3C).

Accuracy differed systematically across auditory conditions and visibility levels (Fig. 3.3B). The mixed-effects model revealed a significant main effect of condition ($p < .001$), with participants performing most accurately in the no-flash baseline (F0B0) and showing progressively lower accuracy as the number of accompanying beeps increased (F1B0 > F1B1 > F1B2). High-visibility locations yielded slightly higher accuracy than low-visibility ones, though this main effect of visibility did not reach significance ($p = .28$), likely due to ceiling-level performance in the F0B0 baseline across all locations. However, visibility strongly interacted with condition (all $p < .001$): accuracy was significantly higher for high-visibility than low-visibility locations when one flash is physically present (F1B0: $\beta = 0.91$, $p < .001$; F1B1: $\beta = 0.66$, $p < .001$; F1B2: $\beta = 0.42$, $p < .001$). No significant main effects or interactions involving group were observed, indicating that low vision and sighted control participants showed similar patterns of performance.

We next examined how flash detection accuracy varied as a function of spatial eccentricity (Fig. 3.3C). The mixed-effects model revealed a significant main effect of condition ($p < .05$): accuracy was highest in the no-flash baseline (F0B0) and declined as the number of beeps increased (F1B2: $\beta = -0.520$, $p < .001$; F1B1: $\beta = -0.220$, $p = .018$; F1B0: $\beta = -0.242$, $p = .01$). Although the main effect of eccentricity was not significant ($p = .78$), several two-way interactions between condition and eccentricity hovered near the threshold for significance (F1B0: $\beta = -0.015$, $p = .089$; F1B1: $\beta = -0.015$, $p = .092$; F1B2: $\beta = -0.013$, $p = .14$).

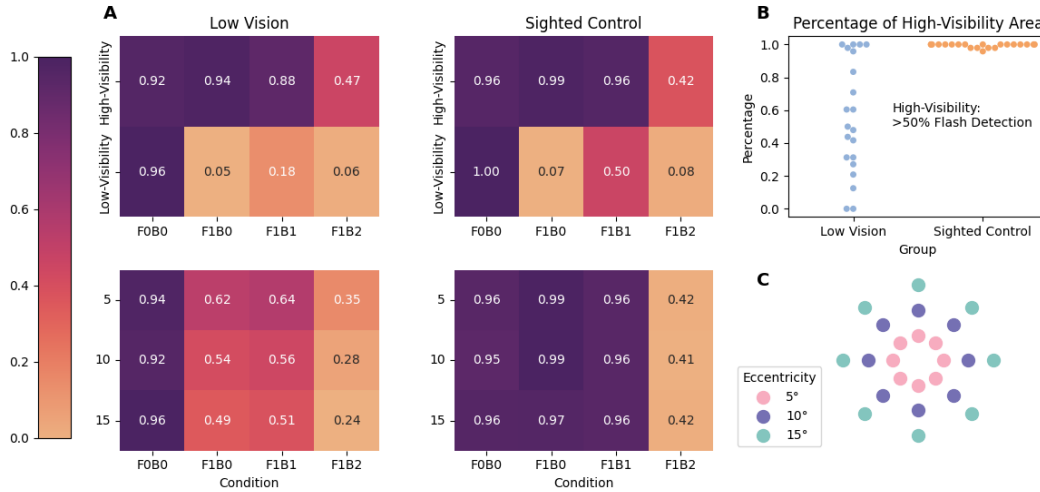


Figure 3.3: Mean accuracy of flash perception across conditions. (A) Heatmaps illustrate the mean accuracy in reporting the veridical number of flashes (0 or 1) across experimental conditions. Locations were categorized into high-visibility (flash-detection accuracy $> 50\%$) and low-visibility ($\leq 50\%$) cohorts based on individual performance in the Visual Flash Detection Task (top panels). Bottom panels display accuracy as a function of stimulus eccentricity (5° , 10° , and 15°). Notably, in the F1B2 (1 flash, 2 beeps) condition, higher accuracy reflects a reduction in the sound-induced flash illusion. Inaccuracy among Low vision participants was split between “misses” (0-flash) and “illusions” (2-flashes), whereas inaccuracy among sighted control participants was driven by a shift toward reporting “illusions” (2-flashes). F0B0 trials yielded the highest overall accuracy, reflecting reliable reports of zero flashes in the absence of stimulation. (B) Proportion of stimulus locations classified as high-visibility for each participant. (C) Spatial layout of stimulus locations, color-coded by eccentricity (5° = pink, 10° = purple, 15° = green), provided for illustrative purposes.

These negative coefficients ($\beta \approx -0.015$) suggest a trend toward decreased accuracy in the periphery under specific multisensory constraints. While these comparisons were likely underpowered to reach standard significance levels at our current sample size ($N = 42$), they suggest that eccentricity’s impact may be mediated by the reliability of the auditory context. Notably, all three-way interactions involving group, condition, and eccentricity were clearly non-significant ($p > .26$), suggesting that the impact of eccentricity did not differ between the sighted control and low vision populations. A modest interaction between group and condition emerged for the F1B0 condition ($\beta = 0.30, p = .024$) and a marginal trend for the F1B1 condition ($\beta = 0.23, p = .079$), indicating slightly higher accuracy for sighted controls in these specific auditory contexts. Beep-detection accuracy was near ceiling across all audiovisual conditions. Because this task was performed only by a subset of sighted control participants, it was not included in the main analyses. The

corresponding heatmap is presented in Appendix S1.

To further investigate why accuracy in the F1B2 (illusion) condition remained low for both groups despite differing mean flash reports, we examined the trial-level response distributions (Supplementary Fig. A.4). For Sighted Controls, inaccuracy in the F1B2 condition was primarily driven by over-reporting (reporting 2 flashes), the hallmark of the Sound-Induced Flash Illusion. In contrast, for the low vision group, inaccuracy was driven by a high frequency of under-reporting (reporting 0 flashes).

As shown in the F1B1 and F1B0 control conditions, low vision participants frequently failed to detect the flash entirely at certain stimulus locations ($\approx 0.5 - 0.6$ accuracy when locations were binned by eccentricity). Consequently, in the F1B2 condition, their mean perceived flashes (reported previously as ~ 1.03) does not represent high accuracy, but rather an average of frequent “misses” (0 flashes) and occasional “illusions” (2 flashes). These distributional data clarify that while both groups show reduced accuracy in the F1B2 condition, the underlying cause for the low vision group is a visual detection floor rather than a heightened susceptibility to the auditory-induced illusion.

Together, these results demonstrate that while both groups exhibited similar patterns of overall accuracy, the underlying drivers of performance were distinct. For sighted controls, errors were primarily illusory, reflecting robust cross-modal integration. In contrast, performance in the low-vision group was fundamentally constrained by a visual detection floor, where frequent “misses” (0 flashes) masked any potential gains in resistance to the illusion. Ultimately, flash numerosity accuracy in this task is determined by an interaction between auditory context and localized stimulus visibility, rather than a global shift in multisensory strategies across the visual field.

Flash Detection vs. Perceived Flash Count (F1B1, F1B2)

We next examined whether individual differences in flash detection accuracy correlated with the strength of sound-induced flash perception. For each participant, eye, and stimulus location, flash detection accuracy in the F1B0 condition was correlated with the mean number of flashes reported in the one-beep (F1B1) and two-beep (F1B2) conditions (Fig. 3.4).

Both groups showed significant positive correlations between flash detection accuracy and perceived flash numerosity for both one-beep (low vision: $r = 0.65$, $p < .001$; sighted control: $r = 0.21$, $p < .001$) and two-beep (low vision: $r = 0.70$,

$p < .001$; sighted control: $r = 0.21$, $p < .001$) conditions. However, visual inspection revealed substantial individual variability, particularly among low vision participants, suggesting that while higher flash detection accuracy generally predicted stronger sound-induced flash enhancement, the strength of this relationship varied markedly across observers.

Together, these results indicate that auditory-driven increases in perceived flash numerosity were present in both groups but expressed with considerable individual variability in magnitude.

Each data point represents a participant's mean flash-detection accuracy (F1B0) and mean number of flashes perceived in a given beep condition. Colors denote participants.

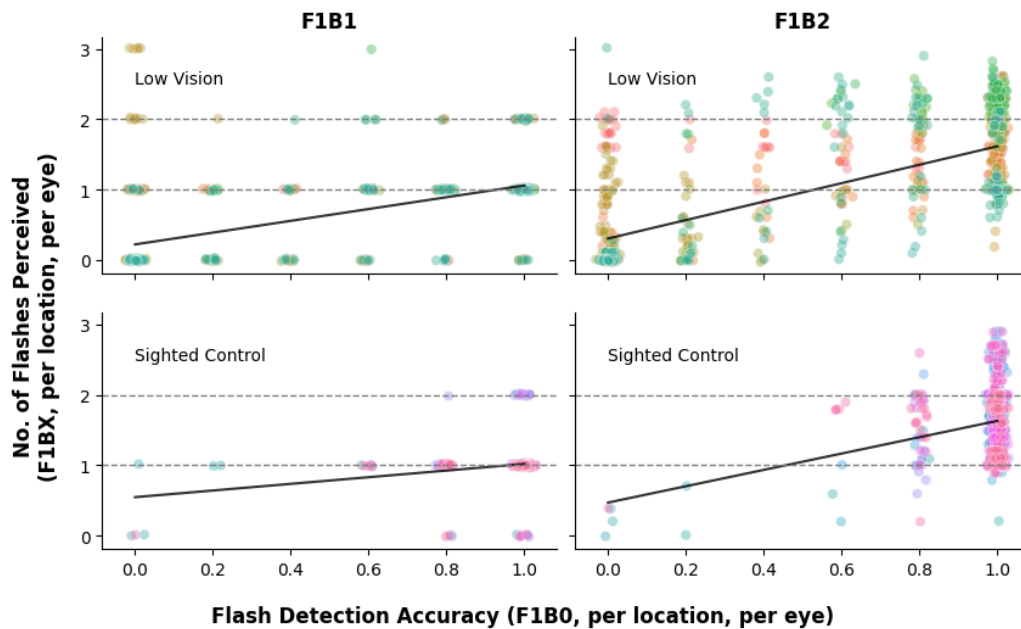


Figure 3.4: **Correlation between flash detection accuracy and perceived flash counts.** Scatter plots show the relationship between flash-detection accuracy (F1B0; Visual Flash Detection Task) and the average number of flashes perceived in the one-beep (F1B1, left) and two-beep (F1B2, right) conditions in the Illusory Double Flash Task. Each data point represents a participant's detection accuracy and their perceived flash count at a specific stimulus location and eye. (48 data points per participant: 24 locations $\times$ 2 eyes) In the F1B1 condition (left), data points represent discrete judgments from individual catch trials presented at each location. In the F1B2 condition (right), data points represent the mean perceived flash count across ten repetitions. Top row: Low Vision; Bottom row: Sighted Control.

Model Fitting

Model Comparison and Selection

To determine which framework best characterized multisensory integration across the visual field, we compared the Bayesian Information Criterion (BIC) scores of four candidate models across all stimulus location subsets (Fig. 3.5A). In both groups, the BCI model consistently outperformed the MLE, forced-fusion (FF), and full-segregation (FS) models, yielding the lowest BIC scores across nearly all spatial partitions. However, model fits were systematically weaker for the low vision group than for sighted controls (Fig. 3.5A; BCI BIC: 524.15 ± 65.25 vs. 281.85 ± 35.99). Similarly, while the BCI model explained all variance in controls ($R^2 = 1.00 \pm 0.00$), its predictive power was slightly reduced in the Low Vision group ($R^2 = 0.97 \pm 0.02$). Across both groups, the BCI framework maintained the superior balance of parsimony and goodness-of-fit.

As trial counts varied significantly across location subsets (Fig. 3.5B), we evaluated model performance using normalized BIC scores (BIC/n ; see Fig. 3.5C). For both groups, the BCI model demonstrated peak performance at high-visibility locations (sighted control: 0.33 ± 0.05 ; low vision: 0.58 ± 0.12). This was corroborated by the R^2 values, where the BCI model explained all variance in the sighted control group ($R^2 = 1.00 \pm 0.00$) and improved predictive power for the low vision group ($R^2 = 0.97 \pm 0.02$) compared to the global fits. In contrast, model fit was substantially reduced for low-visibility locations, with BIC/n scores increasing to 1.40 ± 0.21 for sighted controls and 0.68 ± 0.09 for low vision participants. Correspondingly, R^2 values dropped to 0.73 ± 0.12 and 0.91 ± 0.04 for the sighted control and low vision groups, respectively.

While model fits for sighted controls remained consistently robust ($BIC/n \leq 0.55$) across most spatial partitions, performance for low vision participants degraded when spatial subsets contained heterogeneous visual reliability (e.g., $BIC/n \geq 0.64$ for the All, 5° , 10° , and 15° locations). Notably, two low vision participants had no high-visibility locations and were excluded from subsequent subset analyses. To ensure the most reliable comparison of computational mechanisms, group-level analyses focused on high-visibility locations across the remaining 40 participants (21 Sighted Controls, 19 Low Vision), where model fits were robust in both groups. Because model performance at low-visibility locations was less ideal, particularly for the sighted control group, visibility-level comparisons were restricted to within-group analyses in the Low Vision cohort.

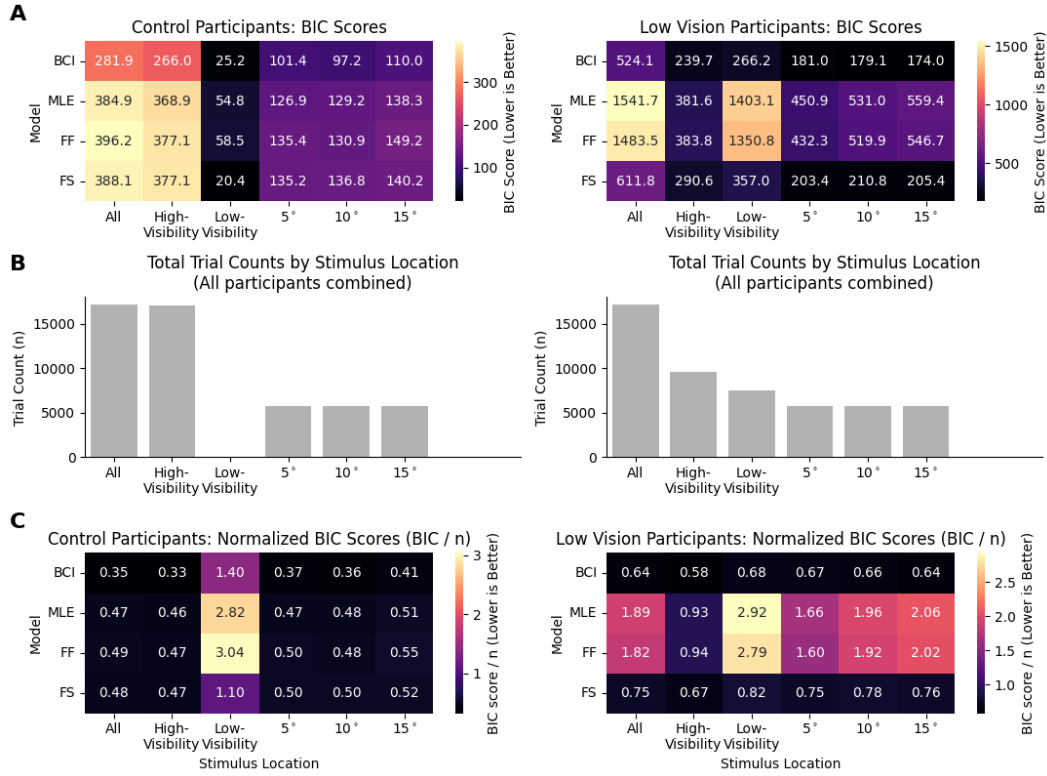


Figure 3.5: Model comparison and selection across stimulus location subsets. BIC evaluates model performance as a function of free parameters and predictive power, lower score means better model performance. **(A)** Raw BIC scores for the four candidate models (BCI, MLE, FF, and FS) across all spatial partitions for Control (left) and low vision (right) groups. Lower BIC scores indicate superior model performance, balancing goodness-of-fit with model parsimony. **(B)** Total trial counts (n) per stimulus location subset, illustrating the data volume used for each model fit. **(C)** Normalized BIC scores (BIC/n), accounting for the varying trial counts across subsets. This normalization reveals that while raw BIC scores may appear lower for low-visibility locations due to smaller sample sizes, the predictive power of the Bayesian framework is optimized at high-visibility locations. Note the consistent superiority of the BCI model (lowest BIC/n) across both groups in high-visibility conditions.

Model Fits and Parameter Estimation

Both groups exhibited the characteristic sound-induced flash illusion (SIFI) in high-visibility locations, as evidenced by increased “2-flash” reports in the F1B2 condition (Fig. 3.6). Qualitatively, sighted controls demonstrated a higher susceptibility to the illusion than the low vision group, more frequently reporting two flashes when a single flash was accompanied by two beeps. Despite these behavioral differences, the BCI framework provided a robust fit for both groups, capturing the response distributions across all multisensory conditions (Fig. 3.6, top rows).

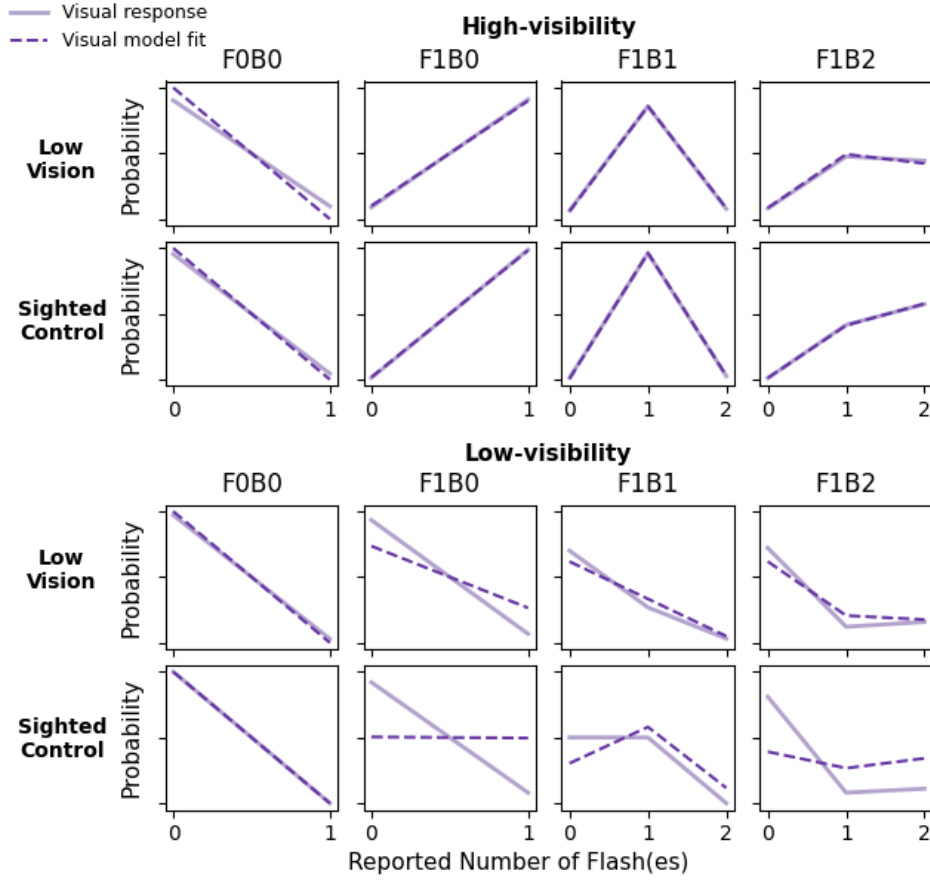


Figure 3.6: **Group-level model fits and parameter distributions for high- and low-visibility locations.** Mean response probabilities (solid lines) and BCI model predictions (dashed lines) for low vision and sighted control groups across stimulus conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps).

Notably, despite the lower R^2 value (0.91 ± 0.04) for model fits at low-visibility locations in the low vision group, the BCI model still captured the overall response pattern reasonably well (Fig. 3.6, bottom rows). In contrast, poorer fits were observed for sighted controls at low-visibility locations ($R^2 = 0.73 \pm 0.12$), likely reflecting insufficient sampling of low-visibility regions rather than a systematic change in inference structure (see Fig. 3.5B). Because model fits for sighted controls in these regions were comparatively unstable, subsequent high- vs. low-visibility analyses focused only on within-group comparisons in the low vision cohort.

Comparison of internal parameter estimates (p_{common} , σ_V , σ_P , and μ_P) revealed limited evidence for group differences at high-visibility locations (Table 3.3). Although p_{common} was significantly lower in the low vision group before correction ($p = 0.01$), the effect was modest after Bonferroni adjustment ($p_{adj} = 0.04$), with

Table 3.3: BCI Parameter Estimates and Group Comparisons Across Visibility Levels

Visibility	Parameter	Sighted Control (Mean $\pm$ SE)	Low Vision (Mean $\pm$ SE)	t	p_{uncorr}	p_{adj}	BF_{10}
High	p_{common}	0.71 ± 0.07	0.42 ± 0.08	-2.69	0.01	0.04	4.80
	μ_P	1.38 ± 0.10	1.24 ± 0.19	-0.66	0.52	1.00	0.37
	σ_P	0.74 ± 0.04	1.10 ± 0.16	2.20	0.04	0.16	1.99
	σ_V	1.40 ± 0.21	1.64 ± 0.22	0.81	0.42	1.00	0.40
Low	p_{common}	0.30 ± 0.13	0.20 ± 0.07	-0.66	0.53	1.00	0.50
	μ_P	0.82 ± 0.44	0.30 ± 0.13	-1.11	0.32	1.00	0.65
	σ_P	1.49 ± 0.62	0.91 ± 0.24	-0.87	0.42	1.00	0.56
	σ_V	0.77 ± 0.07	1.08 ± 0.17	1.72	0.10	0.41	1.11

Note: High visibility comparisons included $N_{Control} = 21$ and $N_{LV} = 19$. Low visibility comparisons included $N_{Control} = 5$ and $N_{LV} = 17$. p_{adj} reflects Bonferroni correction for 4 comparisons within each visibility level. $BF_{10} > 3$ indicates evidence favoring a group difference; $BF_{10} < 1$ indicates evidence favoring the null hypothesis of group invariance.

Bayesian analysis providing moderate evidence for a group difference ($BF_{10} = 4.80$). In contrast, sensory uncertainty (σ_V) and prior mean (μ_P) showed no evidence of group differences ($p_{adj} = 1.00$, $BF_{10} \leq 0.40$), providing support for parameter invariance across groups. Prior uncertainty (σ_P) showed a nominal increase in the low vision group ($p = 0.04$), but this effect did not survive correction ($p_{adj} = 0.16$), and the Bayes Factor remained inconclusive ($BF_{10} = 1.99$). Overall, these findings suggest that under sufficiently visible conditions, the core sensory uncertainty structure of Bayesian causal inference remains largely preserved despite visual impairment, with only limited evidence for altered binding tendency.

Within the low vision group, comparisons between high- and low-visibility locations revealed substantial changes in prior mean estimates (μ_P), which were significantly higher at high-visibility locations (0.30 ± 0.13 vs. 1.24 ± 0.19 , $t = 4.19$, $p_{adj} < 0.001$, $BF_{10} = 61.68$). In contrast, differences in binding tendency (p_{common}) and sensory uncertainty (σ_V) were only nominally significant before correction ($p = 0.05$ for both) and did not survive Bonferroni adjustment ($p_{adj} = 0.18$ and 0.21 , respectively), with Bayes Factors indicating inconclusive evidence ($BF_{10} = 1.37$ and 1.26). Prior uncertainty (σ_P) did not differ between visibility levels ($p_{adj} = 1.00$, $BF_{10} = 0.29$), providing moderate evidence for invariance. Overall, these results suggest that the core causal inference structure remains relatively stable across visibility levels in low vision participants, although reduced visibility may alter

effective prior expectations and introduce noisier sensory estimates. However, these parameter differences should still be interpreted cautiously, as model fit quality decreased at low-visibility locations, indicating that the standard BCI assumptions may capture behavior less reliably under severely degraded visual conditions.

Finally, we evaluated the specific decision strategies employed by participants at high-visibility locations to map their internal causal estimates to a behavioral response (Model Averaging, Probability Matching, or Model Selection). Model Averaging was the dominant strategy for both groups (sighted control: 52.4%; low vision: 47.4%), followed by Probability Matching (sighted control: 33.3%; low vision: 31.6%) and Model Selection (sighted control: 14.3%; low vision: 21.0%).

Model Validation: Identifiability and Recovery

To evaluate the reliability of our computational account, we performed a comprehensive validation of the BCI model’s identifiability and parameter recovery.

First, we examined the identifiability of individual parameters. Under high-visibility context, parameters were entirely independent, with no significant correlations observed between p_{common} , σ_V , μ_P , and σ_P (see Figure A.7). This complete parameter identifiability confirms that the BCI framework successfully dissociates the observer’s internal causal belief from their sensory noise.

Second, we conducted a parameter recovery analysis by simulating synthetic datasets based on our experimental design and participants’ best-fitting parameters. Initial recovery results indicated that all four parameters are well-recovered (p_{common} : $R^2 = 0.94$, σ_V : $R^2 = 0.79$, σ_P : $R^2 = 0.89$, μ_P : $R^2 = 0.90$; see Figure A.8).

3.4 Discussion

We investigated whether low vision alters susceptibility to the Sound-Induced Flash Illusion [70] and whether illusion strength depends on local variations in visual sensitivity. Our results do not support a global “sensory compensation” account, which predicts that chronic visual degradation would lead to a compensatory increase in auditory reliance [11]. Instead, our data suggest that multisensory integration is gated by local sensory reliability.

Low vision participants were not more susceptible to the illusion than sighted controls; in fact, they were less likely to report the categorical “two-flash” percept. This suggests that the “double-flash” illusion is not a default response to visual uncertainty, but rather a sophisticated interaction that requires a minimum threshold

of visual signal to manifest. When visual input is too degraded, the brain may fail to form a sufficiently strong visual representation for the auditory signal to “split”, leading to the reduced susceptibility we observed. Indeed, a similar study investigating factors affecting the susceptibility of SIFI showed that SIFI was robust regardless of the spatial proximity of the audiovisual stimuli, but depended on local visual sensitivity [36].

A key finding of this study is that illusion strength was more robustly predicted by local flash detection accuracy than by spatial eccentricity. While eccentricity-dependent trends approached marginal significance ($p < .10$), they were overshadowed by the direct relationship between a participant’s ability to detect a single flash and their susceptibility to the illusory second flash. This indicates that the “Bayesian brain” does not integrate information based on a fixed spatial map, but rather dynamically weighs cues based on the immediate reliability of the retinal signal at that specific coordinate.

Interestingly, the lack of a robust eccentricity trend was observed not only in the low vision group but also among sighted controls. While classic models of multisensory integration predict increased auditory reliance as visual stimuli move into the periphery [16, 18], our results align more closely with recent findings suggesting that eccentricity alone does not dictate illusion susceptibility [31]. We propose that previously reported eccentricity-dependent effects may serve as a proxy for sensory reliability. The absence of a spatial gradient in our healthy cohort—who exhibited near-ceiling flash detection accuracy across all tested locations—further reinforces our primary conclusion: the sound-induced flash illusion is gated by sensory reliability rather than spatial position. This is supported by a recent computational simulation demonstrating that even when the prior integration tendency (p_{common}) remains constant, the frequency of the illusion increases as the reliability of vision (σ_V) decreases [90]. Conversely, in the low vision group, the reduced susceptibility to the categorical “double-flash” percept suggests that when visual uncertainty is too high, the brain may fail to form a sufficiently stable visual representation for the auditory signal to influence [59]. Consequently, the differences in illusion susceptibility observed in low vision are not an architectural change in integration, but a functional consequence of how the brain manages extreme visual ambiguity.

The exploratory BCI modeling provides the most nuanced insight into this process. When fitted to trials at high-visibility locations, the model produced strong fits for both the low vision and sighted control groups, suggesting that multisensory

sory cue combination and segregation remain broadly consistent with Bayesian inference when visual signals are sufficiently reliable. Interestingly, low vision participants showed a lower p_{common} relative to sighted controls, indicating a weaker prior tendency to infer a common cause across sensory modalities. One possible explanation is that reduced multisensory exposure in daily life due to chronic visual impairment may weaken the development or maintenance of multisensory binding priors [66, 67]. However, this effect should be interpreted cautiously, as only p_{common} survived multiple-comparison correction and the broader parameter pattern remained largely stable across groups.

When comparing fitted BCI parameters between high- and low-visibility locations within the low vision group, we observed a significantly lower μ_P at low-visibility locations, suggesting that participants expected fewer visual events when visual input became unreliable. In contrast, we did not observe statistically reliable changes in σ_V , the parameter theoretically associated with visual sensory uncertainty, nor did we observe strong evidence for changes in binding tendency (p_{common}). Importantly, these parameter estimates should be interpreted with caution. The current experiment was optimized for spatial sampling across many visual field locations rather than exhaustive model fitting, and therefore did not include the full factorial stimulus space of multiple flashes and beeps (e.g., 0, 1, and 2 flashes/beeps). Furthermore, model fit quality decreased at low-visibility locations, suggesting that behavior under severe visual degradation may be noisy or not Bayes-optimal. In these regimes, uncertainty may become underconstrained or partially absorbed into prior-related parameters. Nevertheless, the absence of strong changes in p_{common} across visibility levels suggests that the fundamental causal inference structure remains relatively stable despite degraded vision.

Finally, our behavioral findings refine, rather than contradict, neural evidence showing cross-modal recruitment of occipital cortex in visual impairment [22, 23, 37]. While auditory signals may indeed recruit occipital cortex when some visual function remains, our results demonstrate that such reorganization does not manifest as a global amplification of multisensory interactions. Instead, auditory influence on perceived numerosity is strictly gated by the reliability of local visual input.

Together, these results show that chronic sensory degradation does not automatically lead to compensatory multisensory enhancement. Rather, sound-induced increases in perceived flash count emerge only when visual input is sufficiently reliable to support a stable representation; as vision fails, the brain’s alignment with Bayesian-

optimal predictions becomes systematically weaker and more variable. Ultimately, the Sound-Induced Flash Illusion is a reliability-dependent interaction, and group-level computational comparisons must be interpreted through the lens of local visual accuracy and the resulting limits of model adequacy.

While the present study provides robust evidence for shared multisensory integration strategies between sighted and low vision populations, several limitations warrant consideration.

First, although we observed significant main effects of auditory condition and visual visibility, several interactions involving spatial eccentricity only reached marginal significance ($p < .10$). Specifically, the negative coefficients for the *Condition* $\times$ *Eccentricity* interactions suggest a trend where visual accuracy degrades more sharply in the periphery. As a formal a priori power analysis was not performed for these specific higher-order interactions, and no explicit stopping rule was defined for data collection, it is possible that our sample size ($N = 42$) was underpowered to robustly detect these subtle spatial effects. Future studies with larger cohorts or targeted high-eccentricity paradigms may be necessary to determine if these trends cross the threshold of statistical significance.

Second, our computational approach utilized a Bayesian Causal Inference (BCI) model with certain fixed parameters, such as the auditory noise (σ_A). While this allowed for a more parsimonious model that successfully captured group-level trends, individual differences in sensory processing might be better captured by allowing these parameters to vary. It should be noted that the higher BIC scores observed in the low-visibility locations are partially attributable to the mathematical properties of the likelihood function; as internal noise increases, the maximum achievable likelihood naturally decreases. Future iterations of multisensory models might explore non-Gaussian likelihoods or incorporate additional parameters to account for the specific stochastic nature of clinical vision loss.

Third, while the current study focuses on temporal multisensory integration via the SIFI, future research could employ complementary spatial illusions, such as the Ventriloquist effect [3, 51, 62]. Given that the Ventriloquist illusion typically relies more heavily on visual spatial precision, investigating this effect in low-vision populations would clarify whether the observed preservation of Bayesian integration logic extends from the temporal to the spatial domain.

Furthermore, our low-vision cohort was heterogeneous in diagnosis, the spatial ex-

tent of field loss, and disease duration, likely contributing to the observed variability in both behavior and model fits. While our results suggest a commonality in integration strategies, the specific nature of a visual deficit (e.g., central vs. peripheral loss) may exert unique pressures on the multisensory system that categorical grouping might obscure.

Finally, although eye tracking would have allowed direct verification of fixation stability and retinal locus of stimulation, standard calibration is often infeasible for low vision individuals. At the time of data collection, our laboratory did not yet have accessible calibration tools (e.g., skipped-calibration modes, non-foveal target methods). We therefore relied on task structure and instructions rather than direct gaze measurement. Future work incorporating low-vision-adaptive eye tracking will be crucial for linking local retinal function, fixation behavior, and multisensory integration.

3.5 Conclusion

This study examined how chronic, pathological visual impairment shapes multisensory inference in the sound-induced flash illusion. Contrary to a global “sensory compensation” account, low vision did not lead to a systemic increase in illusion susceptibility. Instead, our results demonstrate that audiovisual interactions are strictly gated by local sensory reliability. Perceived flash numerosity was more robustly predicted by a participant’s ability to detect a single flash than by their clinical group membership or spatial eccentricity. This suggests that the “double-flash” percept is not a default response to visual noise, but a sophisticated interaction that requires a minimum threshold of visual signal to manifest. Our exploratory Bayesian Causal Inference (BCI) modeling further supports the idea that multisensory cue combination remains broadly consistent with Bayesian inference when visual signals are sufficiently reliable. At high-visibility locations, participant behavior closely matched model predictions in both low vision and sighted control groups ($R^2 \geq 0.97$), suggesting that the fundamental computational structure of causal inference remains preserved despite chronic visual impairment. The poor model fits at low-visibility locations likely reflect a combination of noisy behavior that deviates from Bayes-optimal predictions and insufficient data for reliable parameter estimation.

Together, these results demonstrate that degraded vision does not automatically amplify multisensory integration. Instead, the nervous system maintains its fun-

damental rules of inference as long as it is provided with a sufficient threshold of usable sensory input. These findings suggest that potential neural reorganization in low vision does not overwrite the core computational principles of the nervous system. Future work combining low-vision-adaptive eye tracking with models that account for trial-level variability will be essential for a comprehensive account for multisensory integration under extreme, chronic sensory uncertainty.

Chapter 4

MULTISENSORY INFERENCE AT THE BLIND SPOT

The following chapter’s contents are taken and adapted from Chan et al. 2025, with modifications done to fit the dissertation format.

Ailene Y. C. Chan, Noelle R. B. Stiles, Carmel A. Levitan, Daw-An Wu, Armand R. Tanguay Jr., Shinsuke Shimojo (2025). Bayesian causal inference accounts for multisensory filling-in at the blind spot. In: *bioRxiv*
DOI: 10.1101/2024.11.15.623713.

Abstract

We investigated the computational principles of multisensory perception across the physiological blind spot, a region characterized by a total absence of bottom-up retinal input. Using the Audiovisual Rabbit Illusion, we asked three questions: (1) Does multisensory integration persist across the blind spot? (2) Is the illusion amplified by compensatory filling-in? and (3) Does the underlying causal-inference architecture remain stable? Participants reported the number and location of flashes during audiovisual sequences presented either across the blind spot or at a matched control location. Results were analyzed using Bayesian Causal Inference (BCI) modeling and validated via parameter recovery. Robust illusions were observed at both locations. While visual noise (σ_V) was dramatically higher at the blind spot ($BF_{10} > 600$) and priors (σ_P) were significantly expanded, the causal prior (p_{common}) remained functionally stable ($p_{adj} = 0.13$, $BF_{10} = 2.28$). These results demonstrate that the brain sustains coherent multisensory perception through structural gaps by maintaining a stable causal architecture while selectively recalibrating sensory and prior uncertainties.

4.1 Introduction

Perception feels continuous even though sensory inputs are incomplete and unevenly reliable. A major view in perceptual science is that the brain maintains this stability through Bayesian inference, using learned Bayesian priors—internal models of the world—to interpret incoming signals [33]. Some of the clearest evidence for such priors comes from the physiological blind spot, a retinal region with no photoreceptors, where the brain seamlessly “fills in” missing visual information [60]. In

the complete absence of feedforward input, top-down predictions dominate, and the content that is filled in reflects the structure of the brain’s learned internal model rather than external stimulation. Classic accounts emphasize mechanisms such as visual interpolation and predictive coding and thus imply that blind-spot filling-in is supported by strong, highly structured visual priors [25, 61, 84]. Critically, these accounts remain almost exclusively unimodal: they explain how the visual system restores missing visual content, but they do not address how the brain resolves multisensory events that traverse a region with no bottom-up visual signal.

This makes the blind spot an exceptionally powerful test case for multisensory perception: a region with intact spatial and temporal context but no visual input. If perception is guided by stable Bayesian priors about how audiovisual events unfold, do these priors extend to a location where the visual likelihood collapses? One possibility is that multisensory effects should be stronger across the blind spot, reflecting greater reliance on crossmodal cues and a tendency to “fill in” missing information. Another possibility is that filling-in aims to maintain perceptual continuity, in which case multisensory effects should remain unchanged, matching those in fully visible regions.

To answer this, we used the Audiovisual (AV) Rabbit Illusion [74], a spatial-temporal variant of the Sound-Induced Flash Illusion [71]. In the Sound-Induced Flash Illusion, a single flash paired with two beeps leads observers to perceive two flashes, demonstrating that auditory information can alter visual temporal perception. The AV Rabbit Illusion extends this phenomenon across space: when two sequential flashes occurring in different locations are paired with three beeps, participants typically perceive an illusory flash between the two real flashes. Conversely, when three flashes are paired with only two beeps (the “invisible” variant), the middle flash stimulus is perceptually suppressed. This illusion is ideal for probing inference because its percept depends on the brain’s assumed causal structure linking auditory and visual events. We used an established quantitative tool (Bayesian Causal Inference; BCI) to investigate the underlying computations of multisensory perception [4, 38, 68, 69]. If the brain applies the same Bayesian priors across space, illusion strength and model parameters should be identical across the blind spot and a matched control location. If, however, the brain adapts these priors when vision provides no input, multisensory perception should be preserved phenomenologically but supported by different computational signatures—for example, broader visual likelihoods or altered sensory priors.

Participants viewed audiovisual sequences either crossing the blind spot or a visible control region. Although illusion strength was remarkably similar across locations, BCI modeling revealed systematic differences in the underlying computations. Across the blind spot, observers exhibited broader visual likelihoods and flatter sensory priors, indicating that the brain maintains perceptual continuity not by applying the same priors used for visible regions but by flexibly adjusting them when bottom-up visual evidence is absent. Notably, the inferred causal structure—the tendency to bind auditory and visual events—did not differ across locations. Instead, the broader visual likelihoods and flatter priors at the blind spot increase the relative influence of auditory signals on the final percept. These findings suggest that blind-spot filling-in extends beyond vision: the brain engages a form of multisensory filling-in, preserving perceptual continuity through location-specific Bayesian inference even when one modality provides no direct input.

4.2 Methods

Human participants

Twenty participants (10 females; mean age = 26.7 years, range = 19–40 years) took part in the study. All participants had normal or corrected-to-normal vision (20/20) and normal hearing, were naïve to the purpose of the study, and provided written informed consent before participation. Participants were compensated at \$25/hour. All procedures were approved by the institutional review board at the California Institute of Technology (IR19-0926).

Stimuli and conditions

AV Rabbit Task

Visual stimuli were white bars ($0.28^\circ \times 1.2^\circ$ visual angle) presented on a black background (100% contrast) for one frame (16.7 ms at 60 Hz). Auditory stimuli were pure tones (800 Hz, 7 ms). Six audiovisual conditions were tested using the naming convention FXBY, where X is the number of flashes and Y the number of beeps: (i) Unimodal (visual only): F2B0, F3B0, (ii) Bimodal congruent: F2B2, F3B3, (iii) Bimodal incongruent: F2B3 (Illusory AV Rabbit) and F3B2 (Invisible AV Rabbit). Stimulus timing and spatial configurations are illustrated in Fig. 4.1A–B. Stimuli were presented either across the blind spot or a corresponding control location matched in retinal coordinates. On each trial, flashes followed either a horizontal or vertical trajectory. For example, in a F3B0 trial, flashes could appear sequentially left–center–right or top–center–bottom; in a F2B0 trial, only the border positions

(e.g., left–right) were shown.

Blind Spot Mapping

The blind spot of each eye was mapped individually using a flickering circular target (0.5° diameter) alternating between black and white at 7 Hz on a black background. Participants monocularly fixated while moving the target to determine the disappearance and reappearance points along the cardinal directions (left, right, top, bottom). The same locations in the fellow eye's visual field served as control spots, ensuring that stimuli at both positions occupied equivalent spatial coordinates within the visual field.

Rationale for Control Spots

By using the fellow eye's blind spot coordinates as control locations, the experiment compared perceptual performance for identical visual-field positions — once when all flashes were visible (control) and once when one flash fell within the blind spot (partial visual input). For instance, when the left eye was tested, the control spot corresponded to the right eye's blind spot location but was fully visible to the left eye.

Experimental design

Blind spot mapping protocol

Each eye was mapped in three short trials. Participants reported when the flickering target disappeared and reappeared while moving it horizontally or vertically. In trial 1, horizontal borders (left/right) were marked; in trial 2, vertical borders (top/bottom); and in trial 3, horizontal borders were rechecked. Fixation was monitored via an EyeLink 1000 eye tracker. After mapping, the estimated blind spot outline was displayed for participant verification. If borders appeared fuzzy, mapping was accepted; if they appeared sharp, the procedure was repeated.

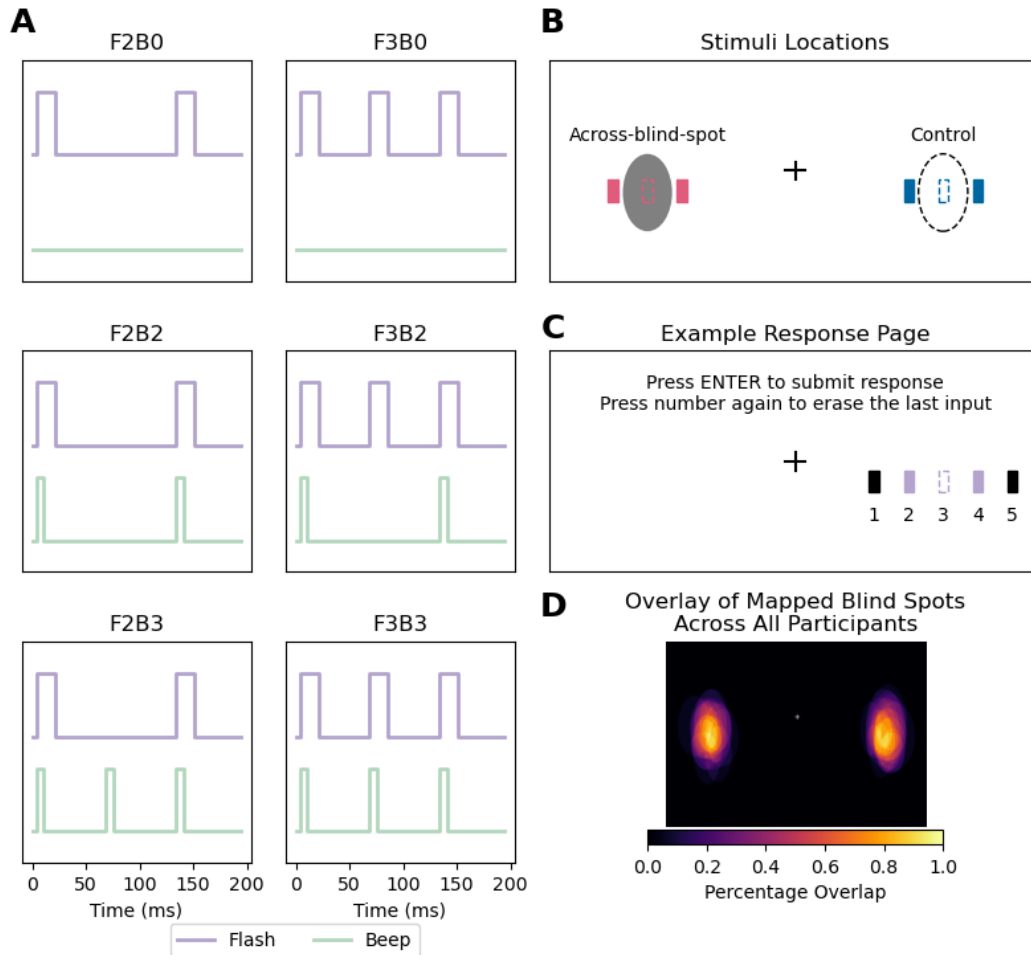


Figure 4.1: **Experimental Design** (A) Schematic of the six experimental conditions, named using an FXBY convention (X = number of flashes, Y = number of beeps). When paired, visual flashes (16.7 ms) and auditory beeps (7 ms) had simultaneous onsets. Key conditions include the **Illusory AV Rabbit (F2B3)**, where an illusory flash is perceived with the second beep, and the **Invisible AV Rabbit (F3B2)**, where the middle flash is perceptually suppressed by the missing second beep. (B) Spatial configuration of stimuli shown for the left eye. Flashes were presented at the center of the blind spot (grey area) or a corresponding control spot (hollow area), as well as 0.5 visual degrees from the cardinal borders of these spots. In $F2BY$ conditions, the two flashes appeared at the solid border points. In $F3BY$ conditions, the first and third flashes appeared at the borders, while the second flash was presented at the center of the blind spot or control spot. For illustration, flashes traveling across the blind spot and control spot are colored blue and red, respectively; all flashes in the experiment were white. (C) The response interface for AV Rabbit Task. Participants reported the number of flashes they perceived by selecting from five equidistant options. Multiple selections were permitted. The colors are for illustrative purposes only. The response interface for the *Additional Flash Localization Task* is shown in Supplementary Video 1. (D) Overlay of the mapped blind spots across participants ($N = 20$). Blind-spot locations and sizes were consistent across individuals (see Fig. S1).

AV Rabbit Task

Participants completed two blocks: one reporting flash perception and one reporting beep perception. Each trial began with 500 ms of stable fixation (verified by eye tracking), followed by the audiovisual sequence and a 500 ms delay before response. In the flash block, participants selected the perceived number and positions of flashes using five equidistant options on screen (Fig. 4.1C). In the beep block, they reported the number of beeps. The relationship between flashes and beeps was never explained.

The design followed a $2 \times 3 \times 4 \times 2$ factorial structure: (i) flashes (2, 3), (ii) beeps (0, 2, 3), (iii) trajectory direction (horizontal/vertical $\times$ 2 directions), (iv) stimulus location (across-blind-spot, control), yielding 48 conditions. Each condition was repeated five times per block (240 trials per block), presented in random order.

Additional Flash Localization Task

A subset of 10 participants returned for additional F2B3 (Illusory Rabbit) localization trials. Each eye was tested separately with 8 trials per eye (4 blind spot, 4 control). Participants, maintaining central fixation, used a crosshair cursor to click perceived flash positions on the screen. The position of the first flash was shown as a reference (see Supplementary Video 1).

Experimental setup

Stimuli were presented using Psychtoolbox 3.0.19 in MATLAB 2024a. Auditory tones (62 dB SPL) were delivered via stereo speakers (Bose Companion 2 Series III) placed beside the monitor. Visual stimuli appeared on a 27-inch LCD display (Dell UltraSharp U2720Q; 60 Hz). Eye position was recorded monocularly at 500 Hz with an EyeLink 1000 (SR Research, Canada) in remote mode. Participants sat 50 cm from the display with their head stabilized on a chin rest and the non-tested eye occluded. The experiment took place in a darkened room. All experimental scripts are publicly available under a CC-BY license [13].

Data analysis

Accuracy

Accuracy was defined as the proportion of trials in which participants correctly reported the true number of flashes and beeps presented (Fig. 4.2A). To test whether the number of beeps influenced flash perception (i.e., the occurrence of AV Rabbit

Illusions), we analyzed accuracy across beep conditions (B0, B2, B3) separately for two- and three-flash trials. For each flash condition and location, accuracy differences across beep levels were assessed using Friedman tests with Bonferroni-corrected Wilcoxon signed-rank post hoc comparisons. Within each condition, accuracy between the blind-spot and control locations was compared using Wilcoxon signed-rank tests (Bonferroni-corrected for multiple comparisons).

Illusion Frequency

To quantify the strength of the two AV rabbit illusions, we computed for each participant and location the proportion of illusion responses (Fig. 4.2B): $P(\text{respond } 3|F2B3)$ for the Illusory AV Rabbit and $P(\text{respond } 2|F3B2)$ for the Invisible AV Rabbit. These proportions were compared between control and across-blind-spot locations using Wilcoxon signed-rank tests (Bonferroni-corrected for multiple comparisons).

Perceived Flash Position

For each AV rabbit illusion, we analyzed the distribution of reported flash positions (1–5) in illusion trials only (Fig. 4.2C). Biases in perceived position were first evaluated with chi-square goodness-of-fit tests against a uniform distribution, and the distributions between control and across-blind-spot locations were compared using chi-square tests of independence.

Signal Detection Theory

To quantify perceptual sensitivity and decision criterion under each condition, we applied signal detection theory (SDT) separately for the Illusory AV Rabbit (3 beeps; B3), Invisible AV Rabbit (2 beeps; B2), and no-beep baseline (0 beeps; B0) contexts (Fig. 4.2D).

In each beep context, trials with three flashes (F3) served as *signal* and trials with two flashes (F2) as *noise*. For each participant and location (across-blind-spot, control), we computed the hit rate $H = P(\text{respond "3"}|F3, B)$ and false alarm rate $FA = P(\text{respond "3"}|F2, B)$, applying a log-linear correction for extreme proportions (Macmillan & Kaplan, 1985). From these, we derived the standard SDT measures of sensitivity (d') and criterion (c) [41]. Sensitivity and criterion were computed as $d' = z(H) - z(FA)$ and $c = -0.5[z(H) + z(FA)]$.

d' reflects the ability to discriminate two vs. three flashes, while c reflects the tendency to report “3”. d' and c differences across beep levels (B0, B2, B3) were assessed using Friedman tests with Bonferroni-corrected Wilcoxon signed-rank post hoc comparisons. Within each beep condition, differences between the blind-spot and control locations were compared using Wilcoxon signed-rank tests (Bonferroni-corrected for multiple comparisons).

Bayesian Causal Inference Model

We modeled participants’ multisensory responses using a Bayesian Causal Inference (BCI) framework, which estimates the probability that auditory and visual cues originate from a common source. Here, we will first describe the BCI model from which we will derive the forced-fusion (FF), full-segregation (FS), and maximum likelihood estimation (MLE) models [29] as alternative models. Details can be found in [38].

The BCI model assumes that each sensory event s_i in the world causes a noisy sensation x_i of the event (Fig. 4.3A). Sensory noise is introduced by drawing x_V and x_A independently from Gaussian distributions centered on the true number of stimuli s_V, s_A with standard deviations σ_V, σ_A shown in Eq. 4.1.

$$x_V \sim N(s_V, \sigma_V), x_A \sim N(s_A, \sigma_A) \quad (4.1)$$

The model also assumes there is a prior bias for sensory information, modeled by a Gaussian distribution centered at μ_P with standard deviation σ_P . The prior distribution is given by Eq. 4.2:

$$x_P \sim N(\mu_P, \sigma_P) \quad (4.2)$$

Common ($C = 1$) and independent ($C = 2$) causes are determined by sampling from a binomial distribution with the causal prior $p(C = 1) = p_{common}$. The posterior probability of the underlying causal structure (unknown to the nervous system) can be inferred by combining the sensory evidence with causal prior according to Bayes rule:

$$p(C = 1 | x_V, x_A) = \frac{p(x_V, x_A | C = 1) p_{common}}{p(x_V, x_A | C = 1) p_{common} + p(x_V, x_A | C = 2) (1 - p_{common})} \quad (4.3)$$

The causal prior quantifies observers' tendency to assume a common cause before stimulus presentation. Post stimulus presentation, sensory evidence from both modalities informs the observers' causal inference via the likelihood term.

If the sensory events can be attributed to a common cause, the estimate of s is a weighted average of both sensory signals and prior for s :

$$\hat{s}_{(V,C=1)} = \hat{s}_{(A,C=1)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_A}{\sigma_A^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}} \quad (4.4)$$

If sensory events are produced by independent causes, the estimate of s is a weighted average of unisensory signals and prior for s :

$$\hat{s}_{(V,C=2)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}, \hat{s}_{(A,C=2)} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}} \quad (4.5)$$

To obtain a final estimate of the number of visual or auditory stimuli, observers' combine estimates under the two causal structures ($C = 1, C = 2$) using various strategies known as “model averaging”, “model selection”, and “probability matching”. Model averaging computes sensory estimates as the weighted sums of maximum likelihood estimates, with posterior probability as weights:

$$\begin{aligned} \hat{s}_V &= p(C = 1|x_V, x_A)\hat{s}_{(V,C=1)} + p(C = 2|x_V, x_A)\hat{s}_{(V,C=2)} \\ \hat{s}_A &= p(C = 1|x_V, x_A)\hat{s}_{(A,C=1)} + p(C = 2|x_V, x_A)\hat{s}_{(A,C=2)} \end{aligned} \quad (4.6)$$

Model selection, as the name suggests, selects sensory estimates based on posterior probability — if $p(C = 1|x_V, x_A) > 0.5$, select $\hat{s}_{C=1}$, select $\hat{s}_{C=2}$ otherwise.

$$\begin{aligned} \hat{s}_V &= \begin{cases} \hat{s}_{V,C=1}, & \text{if } p(C = 1|x_V, x_A) > 0.5 \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_A &= \begin{cases} \hat{s}_{A,C=1}, & \text{if } p(C = 1|x_V, x_A) > 0.5 \\ \hat{s}_{A,C=2}, & \text{otherwise} \end{cases} \end{aligned} \quad (4.7)$$

Probability matching is a stochastic decision making strategy where after calculating posterior probability, draw a random threshold ξ between $[0, 1]$ on each trial, and

select $\hat{s}$ accordingly.

$$\begin{aligned}\hat{s}_V &= \begin{cases} \hat{s}_{V,C=1}, & \text{if } p(C = 1|x_V, x_A) > \xi \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_A &= \begin{cases} \hat{s}_{A,C=1}, & \text{if } p(C = 1|x_V, x_A) > \xi \\ \hat{s}_{A,C=2}, & \text{otherwise} \end{cases}\end{aligned}\tag{4.8}$$

We evaluated model fits by comparing four models: (i) the full Bayesian Causal Inference (BCI) model, (ii) the forced-fusion model (FF; i.e., formally, the BCI model with a fixed $p_{common}=1$), (iii) the full-segregation model (FS; i.e., formally, the BCI model with a fixed $p_{common}=0$), (iv) the maximum-likelihood estimation (MLE) model, which assumes a common cause without prior information (i.e., formally, the BCI model with $p_{common}=1$ and a flat prior, $\sigma_P = 4000$, $\mu_P = 1.5$). An overview of these models is presented in Table 4.1. Because participants differed in their decision strategies (model averaging, model selection, or probability matching), and no single strategy dominated across participants (Supplementary Table B.1), we used each participant’s best-fitting model for parameter estimation. The resulting parameter estimates were qualitatively similar across the three strategies.

Parameter estimation was performed using the BCI Toolbox [89]. Model comparison across strategies used the Bayesian Information Criterion (BIC) to identify the best-fitting model for each participant.

Table 4.1: Overview of models and their free parameters

Model	Free Parameters	#
Bayesian Causal Inference (BCI)	$p_{common}, \mu_P, \sigma_P, \sigma_V, \sigma_A$	5
Forced-fusion (FF)	$\mu_P, \sigma_P, \sigma_V, \sigma_A$	4
Full-segregation (FS)	$\mu_P, \sigma_P, \sigma_V, \sigma_A$	4
Maximum-likelihood estimation (MLE)	σ_V, σ_A	2

4.3 Results

Accuracy

As shown in Fig. 4.2A, for two-flash trials, accuracy significantly decreased when three beeps were presented, confirming that flash perception was significantly influenced by incongruent number of beeps presented (Control: $\chi^2(2) = 29.34$, $p < .001$; Across-blind-spot: $\chi^2(2) = 17.62$, $p < .001$). Similarly, for three-flash

trials, accuracy decreased when only two beeps were presented, confirming that the auditory influence on flash perception (Control: $\chi^2(2) = 31.68, p < .001$; Across-blind-spot: $\chi^2(2) = 32.65, p < .001$).

Post hoc Wilcoxon signed-rank tests confirmed that bimodal incongruent conditions (F2B3, F3B2) yielded the lowest accuracy among all conditions (all Bonferroni-corrected $p < .01$). In contrast, bimodal congruent conditions (F2B2 and F3B3) produced higher accuracy than the corresponding unimodal conditions (F2B0 and F3B0; all Bonferroni-corrected $p < .01$). Notably, for three-flash trials, accuracy in the unimodal F3B0 condition was significantly lower than in the bimodal F3B3 condition across the blind spot, confirming that the middle flash was positioned within the blind spot and that F3B3 functionally resembled the F2B3 (Illusory AV Rabbit) configuration, with participants perceiving the “illusory” (physically present) middle flash.

Beep-report accuracy remained near ceiling across all conditions and locations (mean accuracy > 0.9), confirming that participants reliably discriminated the number of auditory beeps. Because performance showed no meaningful variability across conditions, no further statistical analysis was conducted for beep responses.

Illusion Frequency

Illusion rates were comparable between control and across-blind-spot locations for both the Illusory Audiovisual Rabbit (F2B3) and the Invisible Audiovisual Rabbit (F3B2) conditions (all Bonferroni-corrected $p > .05$), indicating that the illusions occurred with similar strength in both spatial locations (Fig. 4.2B, Fig. B.4).

Perceived Flash Position

For each AV Rabbit illusion (F2B3, F3B2), we analyzed the distribution of reported flash positions (1–5) in illusion trials only (Fig. 4.2C). As expected for the AV Rabbit Illusion, reports were strongly non-uniform across conditions and locations, reflecting the characteristic shift toward the illusory intermediate position. Chi-square goodness-of-fit tests confirmed these patterned distributions (Illusory AV Rabbit (F2B3): Control $\chi^2(4) = 815.27, p < .001$; Across-blind-spot $\chi^2(4) = 805.08, p < .001$; Invisible AV Rabbit (F3B2): Control $\chi^2(4) = 788.68, p < .001$; Across-blind-spot $\chi^2(4) = 779.16, p < .001$).

The distributions of reported positions did not differ significantly between control and across-blind-spot locations (Illusory AV Rabbit (F2B3): $\chi^2(4) = 8.60, p =$

.072; Invisible AV Rabbit (F3B2): $\chi^2(4) = 3.46, p = 0.48$), indicating that both illusions exhibited comparable spatial patterns across locations.

Consistent with the numbered location reports, click-based reports of illusory and last-flash locations collected in the *Additional Flash Localization Task* were centered near the corresponding reference locations for both control and across-blind-spot conditions (Fig. 4.2E). Two-dimensional kernel density maps revealed overlapping clusters of perceived illusory and last-flash positions when aligned to the respective reference points (blind-spot/control center or true last-flash location), suggesting comparable localization precision across locations. The mean distance between perceived illusory flashes and the reference center was 1.38° ($SD = 1.06^\circ$), whereas the mean distance between perceived last flashes and their true physical locations was 1.79° ($SD = 1.40^\circ$), indicating modest variability in spatial accuracy across trials.

Signal Detection Theory

The number of beeps presented significantly modulated sensitivity (d') at control locations ($\chi^2 = 27.1, p < .001$), but not across the blind spot ($\chi^2 = 2.96, p = 0.23$) (Fig. 4.2D). Post hoc Wilcoxon signed-rank tests (Bonferroni-corrected) showed that at control locations, sensitivity was significantly lower for both the two-beep (B2) and three-beep (B3) conditions compared with the visual-only (B0) condition (all $p < .001$), while no difference was observed between the B2 and B3 conditions ($p = 0.81$). At across-blind-spot locations, d' did not differ significantly across beep contexts (all $p > .1$).

On the other hand, the number of beeps presented significantly modulated the response criterion (c) in both control and across-blind-spot locations ($\chi^2 = 40, p < .001$ for each). Post hoc Wilcoxon signed-rank tests (Bonferroni-corrected) confirmed that criterion values were highest for the two-beep condition (B2), lowest for the three-beep condition (B3), and intermediate for the visual-only condition (B0; all $p < .001$). This pattern indicates a systematic response bias, with two beeps biasing reports toward two flashes and three beeps biasing reports toward three flashes.

We compared sensitivity (d') and criterion (c) between control and across-blind-spot locations within each beep condition (B0, B2, B3) using Wilcoxon signed-rank tests with Bonferroni correction. Sensitivity was higher at control than across-blind-spot locations for the unimodal baseline (B0; $W = 0, p < .001$) and Invisible AV Rabbit context (B2; $W = 23, p = .011$). The Illusory AV Rabbit context (B3)

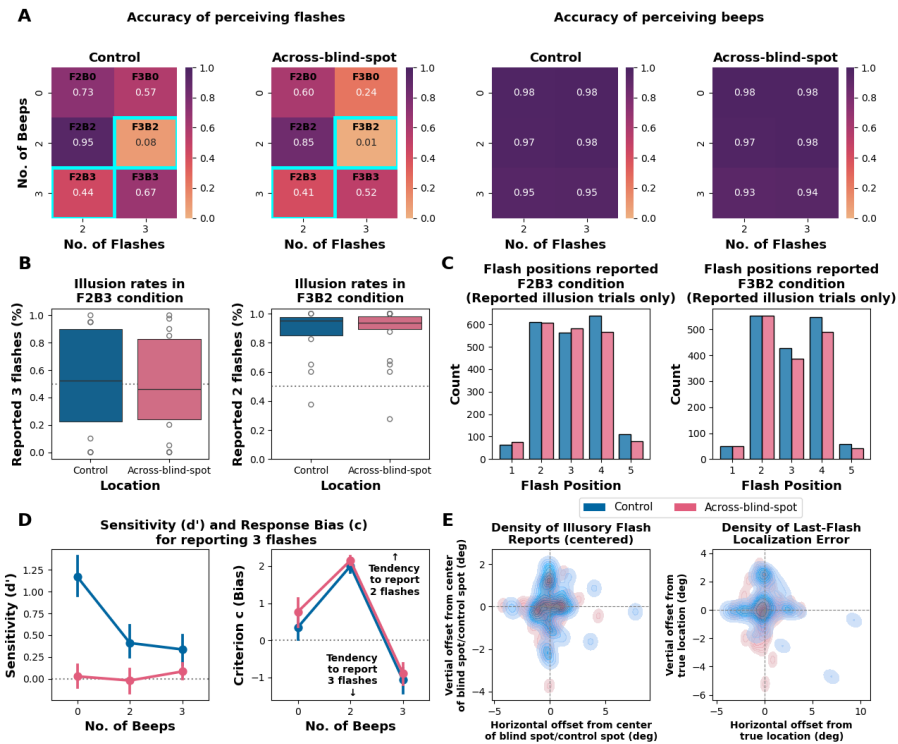


Figure 4.2: Effect of beeps on flash perception. (A) Accuracy for reporting flashes (left) and beeps (right) across control and across-blind-spot stimulus locations. Flash accuracy decreased markedly in conditions where the number of flashes and beeps mismatched (F2B3 and F3B2). Participants also showed reduced accuracy for three-flash stimuli presented across the blind spot in the unimodal (F3B0) condition. Beep-report accuracy remained near ceiling across all conditions. (B) Illusion rates in the Illusory Audiovisual Rabbit (F2B3) and Invisible Audiovisual Rabbit (F3B2) conditions were comparable between stimulus locations. (C) Reported flash positions in the illusion conditions. In F2B3, the extra beep induced an illusory flash between the two real flashes; in F3B2, the missing beep suppressed perception of the middle flash between the first and third flashes. (D) Signal detection results for reporting three flashes. Sensitivity (d') was higher at control than across-blind-spot locations. Response criterion (c) was similar between locations by visual inspection, but it was consistently more conservative across the blind spot, indicating a greater tendency to report “2” across the blind spot. (E) Two-dimensional kernel density maps showing spatial offsets (in degrees of visual angle) of perceived illusory flashes (left) and last-flash reports (right). Offsets are computed relative to the center of the blind-spot/control region for illusory flashes, and relative to the true physical location of the last flash for last-flash localization errors. Horizontal and vertical offsets represent signed deviations along each axis.

showed a similar trend that did not survive correction ($W = 48, p = 0.098$). In contrast, response criterion was consistently more conservative across the blind spot for all beep contexts (B0: $W = 6, p < .001$; B2: $W = 34, p = 0.042$; B3: $W = 28, p = 0.008$), indicating a greater tendency to report “2” across the blind spot.

Although Wilcoxon tests indicated that response criterion (c) was slightly higher across the blind spot than at control locations in each beep condition (all $p < .05$), the mean absolute differences were small in magnitude ($\Delta = 0.18 - 0.37$) and accompanied by substantial overlap of confidence intervals (Fig. 4.2D). These results suggest a very small criterion offset across the blind spot, with no evidence of a qualitatively different decision strategy.

Together, these results show that the blind spot reduces visual discriminability and shifts decision bias away from “3”, whereas the number of beeps primarily modulates response bias (see Friedman analyses), consistent with the AV Rabbit effects.

BCI Modeling Results

Model Comparison

We compared model performance across the Bayesian Causal Inference (BCI; see Fig. 4.3A), forced-fusion, full-segregation, and Maximum Likelihood Estimation (MLE) models using the Bayesian Information Criterion (BIC). For each participant and stimulus location, the model with the lowest BIC score was considered to provide the best trade-off between goodness of fit and model complexity.

Across participants, the BCI model consistently yielded lower BIC scores than forced-fusion, full-segregation, and MLE models (Across-blind-spot: 160.28 ± 12.4 ; Control: 131.47 ± 10.9 ; see Fig. B.5), indicating superior explanatory power. The corresponding forced-fusion (Across-blind-spot: 210.56 ± 20.49 ; Control: 159.67 ± 15.37), full-segregation (Across-blind-spot: 225.85 ± 21.55 ; Control: 191.37 ± 13.70), and MLE (Across-blind-spot: 245.33 ± 23.71 ; Control: 183.02 ± 19.02) models fit the data substantially worse.

BCI Model Estimated Parameters

We fitted a Bayesian Causal Inference model with five free parameters for each participant, stimulus location (across-blind-spot and control), and decision strategy (model averaging, model selection, and probability matching). For each participant

and location, the best-fitting strategy was identified by comparing Bayesian Information Criterion (BIC) values across the three models. Table 4.2 summarizes the estimated parameters from these best-fitting models (see Fig. 4.3B for the visualization of group-average fitted parameters at across-blind-spot). Because participants differed in their preferred strategies and no single strategy consistently provided the best fit (Supplementary Table B.1), parameter estimates were taken from each participant's individually best-fitting model. Overall, the estimated parameters were qualitatively similar across strategies.

Goodness of fit was quantified using the coefficient of determination (R^2) between the model-estimated numerical responses and the participant-reported numerical responses (i.e., perceived number of beeps and flashes). The best BCI models demonstrated excellent performance, achieving a mean R^2 of 0.99 across all 20 participants (see Fig. 4.3C for group-average model fitting results).

Table 4.2: BCI Model Parameter Estimates and Statistical Comparisons (N=20)

Parameter	Across Blind Spot (Mean $\pm$ SE)	Control (Mean $\pm$ SE)	t	p_{uncorr}	p_{adj}	BF_{10}
p_{common}	0.50 $\pm$ 0.06	0.58 $\pm$ 0.07	-2.40	0.03	0.13	2.28
σ_P	1.19 $\pm$ 0.13	0.84 $\pm$ 0.07	3.31	<.01	0.02	12.00
μ_P	1.90 $\pm$ 0.14	2.33 $\pm$ 0.16	-3.25	<.01	0.02	10.66
σ_V	2.32 $\pm$ 0.12	1.32 $\pm$ 0.15	5.30	<.001	<.001	619.68
σ_A	0.21 $\pm$ 0.02	0.19 $\pm$ 0.01	1.05	0.31	1.00	0.38

Note: p_{adj} reflects Bonferroni correction for 5 comparisons. $BF_{10} > 3$ indicates moderate evidence for a difference; $BF_{10} < 0.33$ indicates moderate evidence for the null.

Statistical comparisons of parameter estimates between the across-blind-spot and control locations revealed significant differences in visual and numerical prior parameters (Table 4.2). Visual noise was substantially higher at the blind spot ($\sigma_V : t(19) = 5.30, p < .001, BF_{10} = 619.68$), and numerical priors were found to be both wider ($\sigma_P : t(19) = 3.31, p_{adj} = 0.02, BF_{10} = 12.00$) and shifted ($\mu_P : t(19) = -3.25, p_{adj} = 0.02, BF_{10} = 10.66$). Notably, while the causal prior p_{common} showed a nominal uncorrected difference ($t(19) = -2.40, p = 0.03$), this effect did not survive Bonferroni correction ($p_{adj} = 0.13$) and Bayesian analysis yielded only anecdotal evidence for a shift ($BF_{10} = 2.28$). In contrast, auditory precision remained stable across locations, with Bayesian evidence favoring the null hypothesis ($\sigma_A : t(19) = 1.05, p_{adj} = 1.00, BF_{10} = 0.38$).

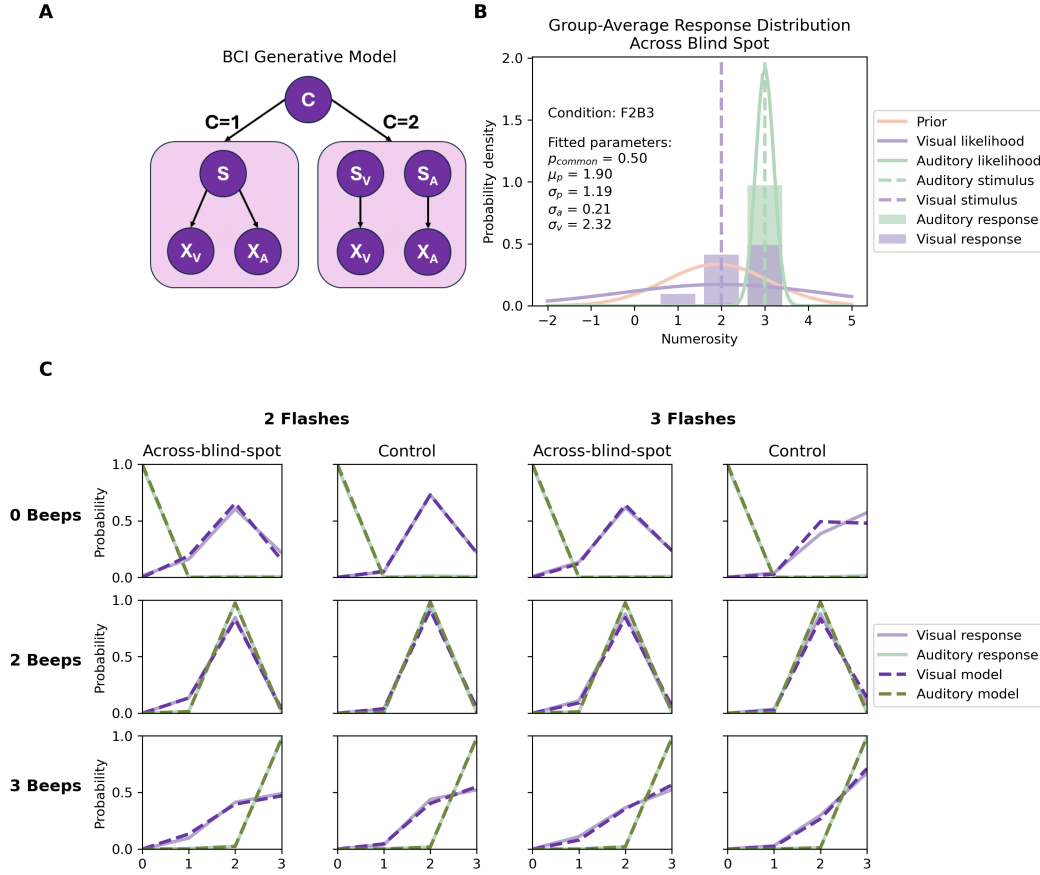


Figure 4.3: The generative model of the Bayesian Causal Inference model and group-average model fitting results. (A) The BCI model where multisensory cues either share a common cause ($C=1$) or independent causes ($C=2$). If cues share a common cause, the final percept s is derived by cues from both modalities, i.e., x_V and x_A , each perturbed by noises with width σ_V and σ_A , respectively. If cues are independent, final percepts s_V and s_A are derived independently from x_V and x_A . (B) The BCI model infers the causal structure by combining sensory likelihood and prior. Prior parameters include prior stimulus expectation and p_{common} (a prior expectation of a common cause). Sensory likelihood represents the sensory input (x_V, x_A) perturbed noise (σ_V, σ_A). Here, we have the group-average fitted parameters across-blind-spot and the predicted probability of auditory and visual responses. Since auditory likelihood is narrow compared to visual likelihood, the final percept s is heavily influenced by the auditory stimulus. (C) Group average model fitting results. Solid purple and green lines correspond to the visual and auditory response data, dashed dark purple and dark green lines correspond to the model predictions for visual and auditory responses. Given the robust BCI fits (mean $R^2 = 0.99$), we see almost a complete overlap of response data and model predictions.

Model Validation and Parameter Recovery

To ensure the reliability of our parameter estimates and test for model identifiability, we performed a parameter recovery analysis. We simulated synthetic datasets using

the experimental design of the current study and parameter values drawn from the range of our participants’ best-fitting models. We then applied the same fitting procedure to these synthetic data to determine if the true (generative) parameters could be accurately recovered. As shown in Figure B.6, the model demonstrated robust recovery performance. We observed strong correlations between true and recovered values for the primary parameters of interest: visual noise (σ_V , $R^2 = 0.83$), numerical prior width (σ_P , $R^2 = 0.94$), and auditory noise (σ_A , $R^2 = 0.97$). While the causal prior (p_{common}) and numerical prior mean (μ_P) also showed excellent recovery ($R^2 = 0.95$ and $R^2 = 0.99$, respectively), they exhibited the coupling identified in our correlation analysis (see Fig. B.7; ($r = 0.66$)). This may reflect an inherent identifiability constraint in the BCI model under low-density sampling, likely representing a combined numerical-causal prior cluster. The remaining parameters (σ_V , σ_A , σ_P) were successfully isolated by the model. These results confirm that our experimental paradigm provides sufficient data density to uniquely identify the key sensory and prior parameters, providing high confidence in the observed differences between the blind-spot and control locations.

Together, these findings indicate that the “filling-in” process at the blind spot is characterized by a dramatic increase in visual uncertainty (σ_V) and a systematic expansion of internal numerical priors (σ_P). While we observed a nominal reduction in the binding prior (p_{common}) at the blind-spot location, the associated Bayesian evidence was anecdotal ($BF_{10} = 2.28$), and this parameter was found to be coupled with the numerical prior mean (μ_P). This suggests that while sensory noise and numerical precision are significantly altered by the absence of bottom-up input, the underlying causal architecture remains functionally stable, with any potential shifts in binding tendency being intrinsically linked to the brain’s top-down reconstructive numerical expectations.

4.4 Discussion

In this work, we asked whether multisensory filling-in extends into the physiological blind spot, and if so, how this process is computationally achieved. We asked three questions: (i) Does audiovisual integration produce the Audiovisual Rabbit Illusion [74] across the blind spot? (ii) If so, is the illusion enhanced because the brain has a stronger tendency to fill in missing information where vision is absent? (iii) Even if behavior appears similar across regions, are the underlying computational mechanisms the same or fundamentally altered in the blind spot? Our findings address these questions in turn.

Does multisensory filling-in occur across the blind spot?

Yes. Participants experienced robust Audiovisual Rabbit Illusions in both the classic (F2B3) and invisible (F3B2) variants across the blind spot (Fig. 4.2A). This demonstrates that audiovisual binding does not require direct visual input from the interpolated region. Instead, the percept appears to depend on internal generative models that enforce spatial and temporal continuity. Prior work on filling-in has largely emphasized unimodal mechanisms [25, 35], or compensatory enhancements in long-term blindness [11, 37]. Our results extend these insights to multisensory perception in sighted observers: the brain treats the blind spot as part of a continuous visual scene and incorporates auditory structure to construct a percept that is not directly supported by retinal input.

Localization data reinforce this conclusion. Number-based reports showed that perceived flashes – illusory or real – clustered near their expected spatial positions (Fig. 4.2C). Click-based reports, collected from a smaller subset with limited trials, showed a qualitatively similar pattern (Fig. 4.2E), though these results should be interpreted with caution given the substantially smaller dataset. Mean localization errors remained within a few degrees, suggesting that spatial interpolation at the blind spot preserves the coordinate structure necessary for multisensory alignment.

Is the illusion enhanced across the blind spot due to stronger filling-in?

No. Contrary to the possibility that missing visual information might amplify reliance on auditory cues, illusion strength was statistically indistinguishable across blind-spot and control regions (Fig. 4.2B). This suggests that the brain does not default to a stronger common-cause assumption when visual evidence is absent. In other words, crossmodal signals do not automatically dominate at the blind spot.

This finding constrains interpretations of sensory compensation. Although compensatory enhancements have been documented in long-term or congenital visual deprivation [56], our results show that compensation at the blind spot is not expressed as increased multisensory integration. Instead, the blind spot is integrated using the same overall inference structure as visible regions.

Are the underlying computational mechanisms preserved or altered?

Here, our computational modeling reveals a nuanced dissociation: while the fundamental Bayesian Causal Inference (BCI) framework remains the most robust account of behavior in both regions, the internal representation of uncertainty and prior expectations undergoes systematic recalibration.

BCI provided the best quantitative account of behavior at both locations, achieving an omnibus $R^2 = 0.99$ and outperforming forced-fusion, full-segregation, and MLE models. This replicates prior demonstrations that observers dynamically arbitrate between common- and separate-cause interpretations [38, 69]. Crucially, our parameter recovery analysis confirmed that even with the unique constraints of the blind-spot geometry, our model could accurately recover generative parameters (all $r > 0.83$), providing high confidence in the following observed dissociations (Table 4.2):

1. **Stability of Causal Architecture:** We found that the tendency to infer a common cause (p_{common}) was functionally stable across regions. While we observed a nominal decrease in p_{common} at the blind spot, this effect did not survive correction for multiple comparisons ($p_{adj} = 0.13$) and yielded only anecdotal Bayesian evidence ($BF_{10} = 2.28$). This suggests that the brain does not fundamentally alter its binding tendency simply because sensory evidence is reconstructed rather than received.
2. **Coupling of Causal and Numerical Priors:** A key finding was the strong correlation between the causal prior (p_{common}) and the spatial prior mean (μ_P) ($r = 0.66$, $BF_{10} > 5000$). This coupling suggests that in the absence of bottom-up visual input, the brain’s belief in a common source becomes structurally tied to its top-down prior expectations. This dependency represents an adaptive strategy: when sensory evidence is missing, the system relies on a more integrated prior “cluster” to maintain perceptual stability.
3. **Recalibration of Sensory and Prior Uncertainty:** While the architecture remained stable, the parameters defining the sensory input were significantly altered. Visual noise (σ_V) was dramatically higher at the blind spot ($BF_{10} = 619.68$), representing the primary perceptual consequence of absence of photoreceptors. Furthermore, numerical priors were both wider (σ_P , $BF_{10} = 12.00$) and shifted (μ_P , $BF_{10} = 10.66$). This indicates that the brain accounts for the lack of retinal input by loosening its prior expectations, allowing for a broader range of possible interpolated percepts.

These findings align with the proposal that binding tendencies are stable within specific tasks [48]. However, our results extend this by showing that even when

a modality is muted by physiological gaps, the system maintains its causal framework by selectively adjusting uncertainty parameters. This separation between architecture (preserved) and parameterization (recalibrated) shows that multisensory inference is resilient to structural gaps in input.

Limitations and Future Directions

We acknowledge that the comparison between the blind spot and the fellow-eye control involves inherent asymmetries in nasal and temporal retinal processing. While these low-level differences exist, they are unlikely to account for the specific pattern of multisensory rabbit effects observed here because stimuli were presented in both directions randomly (nasal-temporal and temporal-nasal). Furthermore, while the high correlation between p_{common} and μ_P reflects a constraint in model identifiability under restricted data density, our parameter recovery results confirm that the core sensory parameters (visual and auditory noise; σ_V , σ_A) are uniquely and robustly identified.

Overall significance

Together, these results characterize how the brain fills in the blind spot in multisensory contexts. The perceptual system maintains continuity not by increasing binding tendencies, but by recalibrating its internal model of noise and prior expectations on the number of sensory events. Multisensory perception persists through a sophisticated probabilistic reconstruction, revealing a system that sustains a coherent view of the world by relying on its structural architecture even when the local sensory data is missing.

4.5 Conclusion

This work demonstrates that multisensory filling-in at the blind spot is not a simple heuristic “patching” of missing data, but a sophisticated process of probabilistic reconstruction. By applying a Bayesian Causal Inference framework, we have shown that the brain maintains a remarkably stable computational architecture even in the complete absence of bottom-up visual input. The persistence of the Audiovisual Rabbit Illusion across the blind spot confirms that multisensory integration is a top-down, generative process that enforces continuity regardless of retinal gaps.

Our modeling results provide a high-resolution map of how this is achieved. While the binding tendency (causal priors) remains relatively invariant, the system adaptively recalibrates its parameters, specifically by increasing visual uncertainty and

widening priors, to reflect the lack of direct sensory evidence. The discovery of a functional coupling between causal and numerical priors further suggests that in the absence of input, the brain's internal model of the world becomes more integrated, relying on a unified top-down cluster to resolve ambiguity.

Ultimately, these findings reveal the resilience of the human perceptual system. We do not perceive gaps in our multisensory world because the brain possesses a robust internal map of causal structure that can tolerate local sensory failure by adjusting its confidence, rather than abandoning its core principles of inference.

MULTISENSORY INFERENCE IN VIRTUAL ENVIRONMENT

Abstract

This study characterizes visual–tactile integration within a fully immersive 360° environment. By systematically manipulating spatial discrepancy and visual reliability, we applied a Bayesian Causal Inference (BCI) framework to examine how the brain resolves sensory binding across the full azimuth. A central finding is that under conditions of extreme visual noise, the descriptive performance of BCI degrades: formal model comparison (Bayesian Information Criterion) favored a simpler Maximum Likelihood Estimation (MLE) fusion model. Importantly, this does not indicate that MLE provides a better account of behavior. In fact, MLE exhibited poor absolute fit and failed to capture behavioral variability, whereas BCI continued to track behavior.

Despite reduced goodness-of-fit, BCI model metrics remained systematically related to behavior ($r \approx -0.72$), whereas simpler models did not, indicating that BCI retains sensitivity to task-relevant structure even when its descriptive accuracy is limited. This dissociation highlights a gap between statistical model selection and explanatory validity, suggesting that degraded sensory input constrains the effective expression or recoverability of Bayesian computations rather than invalidating their underlying structure. Using BCI-derived parameters, we further show that tactile sensitivity varies across regions of the torso, while the binding tendency for integrating visual–tactile cues remains invariant across space.

Together, these findings suggest that multisensory integration is shaped by sensory reliability: while the brain may retain a stable inferential architecture, its effective operation becomes increasingly difficult to capture under high uncertainty. More broadly, this work positions BCI as a powerful but reliability-dependent framework, with important implications for both theory and the design of sensory substitution and navigation systems.

5.1 Introduction

Navigating complex, 360-degree environments requires the human nervous system to continuously integrate noisy and often conflicting sensory signals. Virtual reality

(VR) allows researchers to investigate this highly dynamic process while maintaining a high degree of experimental control [42]. Current head-mounted displays (HMDs) offer a large binocular field of view for natural interaction, but selecting peripheral objects outside this view requires extensive visual search through head rotations, which can negatively impact performance and subject observers to simulator sickness [5]. Tactile displays—particularly vibrotactile belts worn around the abdomen—provide a highly effective, complementary channel for omnidirectional spatial cueing, as the torso possesses a natural egocentric mapping to external directions [19, 20, 77, 79]. When multiple sensory cues are available, the human nervous system typically integrates visual and tactile information in a statistically optimal fashion [29], a process that is highly effectively modeled by Bayesian Causal Inference (BCI) [38, 69, 88].

However, the mechanisms governing this multisensory integration are underexplored in naturalistic environments. While BCI has been studied extensively within the visual-auditory domain using spatial localization tasks, these investigations have almost exclusively been restricted to 2D planes or limited frontal fields of view [4, 10, 48–50, 63]. Computational models of visual-tactile integration are even less common; existing studies are either similarly restricted to frontal fields of view [81] or focused on the numerosity domain [87]. Although vibrotactile feedback and spatial guidance have been studied extensively in electrical engineering and human factors fields to evaluate their practical effectiveness [24, 40, 58, 78], these investigations rarely address the underlying probabilistic mechanisms of sensory integration. Consequently, it remains unclear how the nervous system computationally resolves visual-tactile integration under varying environmental contexts, such as when visual information is highly unreliable or degraded.

Understanding these computational mechanisms under visual uncertainty has broad and critical applications. For individuals with low vision or blindness, elucidating how the brain weighs degraded visual input against tactile cues can directly inform the design and optimization of sensory substitution device [27, 53, 73]. Even for sighted individuals, these mechanisms are highly relevant in demanding practical scenarios. For example, during combat operations, soldiers face extreme sensory demands in environments where visual and auditory channels are frequently compromised by hazards like dust, smoke, and gunfire. Relying on conventional visual or auditory interfaces in these chaotic settings can overwhelm a soldier’s cognitive capacity, ultimately impairing their operational effectiveness. To overcome these

bottlenecks, spatial tactile displays provide a reliable alternative for transmitting critical directional data, bypassing overburdened sensory pathways without adding to the user’s mental workload [28, 42].

To address these gaps, the present study investigates visual-tactile integration under different visual contexts within a 360-degree space. Participants performed a directional targeting task using a custom-built vibrotactile belt featuring eight eccentric-rotating-mass (ERM) tactors integrated with a VR HMD. We systematically manipulated the spatial discrepancy between visual and tactile targets, as well as the reliability of the visual signal using an adjustable alpha mask over the main camera. To explicitly link internal sensory processing with physical motor execution, we evaluated participant behavior using a dual analytical approach: applying a BCI framework to model the probabilistic integration of sensory cues, and extracting discrete kinematic features to decode the underlying continuous motor control strategies. By forcing sighted observers into high-uncertainty regimes, this study aims to clarify whether the rules of sensory integration are spatially invariant and how visual degradation dictates the physical execution of movements in a 360-degree space.

5.2 Methods

Human participants

Twenty-five participants (15 females; range = 18-53 years, mean = 30.1, SD = 7.5) took part in the study. All participants had normal or corrected-to-normal (20/20) vision. With the exception of one author ($N = 1$), all participants were naïve to the purpose of the study. Written informed consent was obtained from all participants prior to the session. Participants were compensated at a rate of \$25/hour. All procedures were approved by the Institutional Review Board at the California Institute of Technology (Protocol IR24-0926).

Materials and apparatus

Vibrotactile Belt

Directional tactile cues were delivered via a custom-built vibrotactile belt constructed from a loop-style inner duty belt (Fig. 5.1A). The system featured eight miniature eccentric-rotating-mass (ERM) tactors (Pico Vibe 9 mm, 25 mm type; model 307–103; Precision Microdrives, London, UK) controlled by an Arduino Nano 33 BLE Sense Rev2. The tactors were driven by a custom printed circuit

board (PCB) designed by D. Wagenaar and A. Chan and assembled in-house (Fig. 5.1B).

Each tactor was housed in an individual felt pocket with Velcro backing, allowing the tactors to be repositioned easily to accommodate different waist sizes. To prevent rotation of tactors inside the felt pocket, which would compromise wire integrity, D. Wagenaar designed and 3D-printed tactor holders to restrict tactor rotations. Tactors were aligned with the participant's cardinal (N, E, S, W) and collateral (NE, NW, SE, SW) points relative to the torso. Detailed circuit schematics and PCB layouts are provided in Fig. C.1.

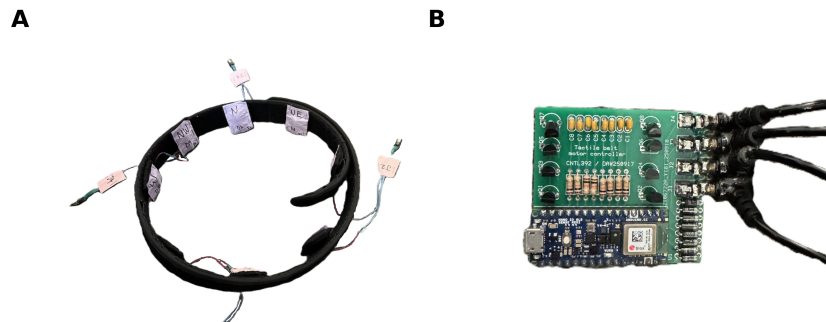


Figure 5.1: **Vibrotactile Belt Prototype and Control Hardware.** (A) The wearable belt assembly featuring eight tactors distributed at cardinal and collateral positions. (B) The custom PCB interface integrated with an Arduino Nano 33 BLE Sense Rev2 for signal processing and wireless communication.

Virtual Reality Headset

Visual stimuli were presented using a Meta Quest 2 head-mounted display (HMD). Participants interacted with the environment using a single right-hand VR controller to aim and “shoot” at targets.

Software

The experimental task was developed in Unity (Unity Technologies) using custom C# scripts. The virtual environment placed participants at the center of an open space (Fig. 5.2A) where they scanned the horizon for Unidentified Flying Objects (UFOs; Fig. 5.2B) to engage with a virtual laser gun (Fig. 5.2C). 3D assets were

sourced from Sketchfab under a standard license. Representative gameplay and trial procedures are illustrated in Supplementary Video 2.

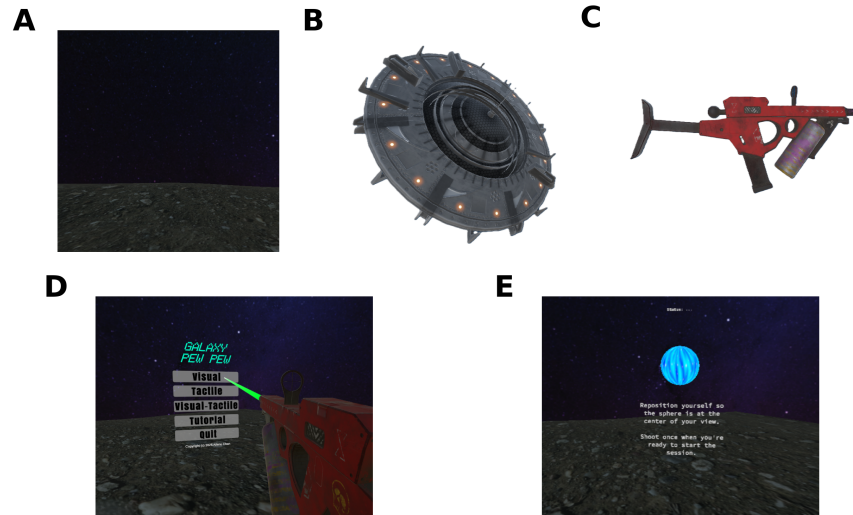


Figure 5.2: **Visual assets and integrated environment.** (A) Base terrain showing surface texture and starfield background. (B) Unidentified Flying Objects (UFO) model. (C) Laser gun used for player interaction. (D) Final composite scene of the main menu interface demonstrating the integration of all assets within the game engine. (E) Calibration scene.

Communication between the Unity application and the Arduino-based belt was established via Bluetooth Low Energy (BLE) using the *Bluetooth LE for iOS, tvOS and Android plugin* (Shatalmic, LLC). The software managed trial randomization, target positioning, and tactor activation. It recorded behavioral data at 60 Hz, including shooting accuracy (angle shot, angular error), reaction time, and continuous heading (yaw) for both the HMD and the handheld controller.

Design and procedure

The experimental task was a spatial shooting paradigm implemented in a fully immersive virtual environment. Participants were positioned at the center of an open 360° space. At the beginning of each trial, a visual target appeared at a random direction at eye level. Participants were instructed to remain seated but were free to move their upper bodies to scan the environment, locate the target, and execute one shot.

In tactile-only blocks, no visual target was presented. Instead, participants were instructed to orient and shoot in the direction indicated by vibrotactile stimulation

delivered to the torso. On each trial, the stimulus (visual or tactile) was presented for 3 seconds, after which participants were required to respond regardless of whether the target had been successfully localized.

Auditory cues were included for environmental ambiance and trial timing. Specifically, an alarm sound signaled the onset of each trial when the target appeared. However, all sounds were non-spatialized (2D) and did not convey directional information.

The session consisted of six blocks: two unimodal baseline blocks followed by four multisensory blocks. Participants always completed the unimodal Visual (V) and Tactile (T) blocks first to establish baseline performance. At the start of each block, participants selected the corresponding block type (V, T, or VT) or a tutorial from the Main Menu Scene (Fig. 5.2D). This was followed immediately by a Calibration Scene (Fig. 5.2E), allowing participants to physically reposition themselves so their “true North” aligned with the in-game orientation. Once calibrated, participants entered an Instruction Scene and initiated the experiment when ready. A total of eight directions were tested: cardinal (N, E, S, W) and collateral (NE, NW, SE, SW). For analysis, these eight directions were further categorized into three regions: front (N, NE, NW), side (E, W), and back (S, SE, SW) (Fig. 5.3A).

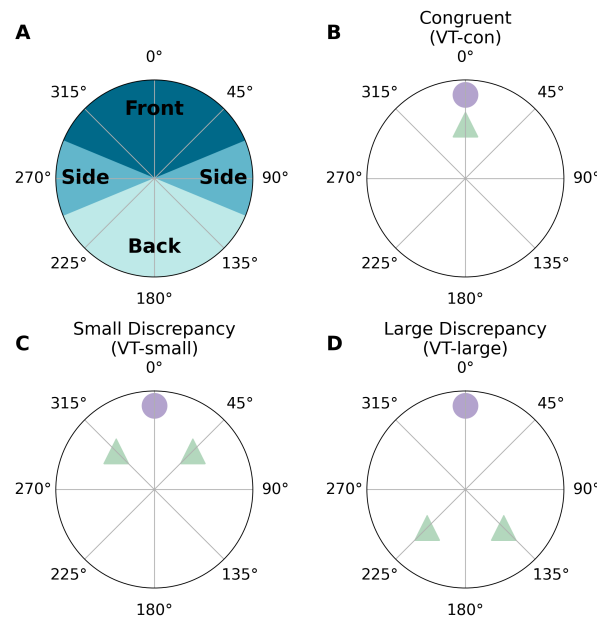


Figure 5.3: **Visual and Tactile Stimulus Directions** A: Stimulus directions categorized into three regions: front, side, and back. B-D: Visual (purple circle) and possible tactile (green triangle) stimuli directions in VT-con, VT-small, and VT-large conditions

The subsequent multisensory Visual-Tactile (VT) blocks were grouped by required response modality (Visual or Tactile). To mitigate fatigue, each response modality condition was split into two blocks. Within VT blocks, trials were categorized into three spatial subconditions: (1) **Congruent** (VT-congruent): 0° spatial offset between visual and tactile stimuli (Fig. 5.3B), (2) **Small Discrepancy** (VT-small): $\pm 45^\circ$ offset (e.g., a 0° visual target paired with a NW or NE tactor; Fig. 5.3C), (3) **Large Discrepancy** (VT-large): $\pm 135^\circ$ offset (e.g., a 0° visual target paired with a SW or SE tactor; Fig. 5.3D). For both discrepancy conditions, the direction of the offset (positive or negative) was randomized across trials within each block.

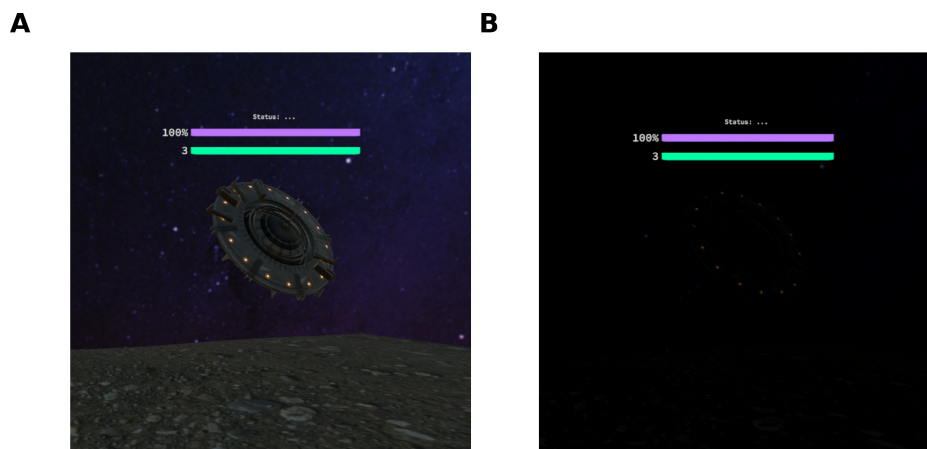


Figure 5.4: **Visual Reliability Levels.** (A) High Reliability ($\alpha = 0$). (B) Low Reliability ($\alpha = 245$).

Visual reliability manipulation was achieved by overlaying a spherical mask over the main camera (participants' first person point of view) and adjusting its alpha (opacity) in Unity: **High Reliability** ($\alpha = 0$) and **Low Reliability** ($\alpha = 245$; max 255) (Fig. 5.4). The alpha value for the low reliability condition was selected to maximize the contrast between conditions within the technical constraints of the headset. Although higher opacity values were tested, $\alpha = 245$ represented the highest level at which the visual target remained faintly visible, avoiding complete disappearance due to display resolution limits.

The unimodal V block followed a $2 \times 8 \times 5$ factorial design (Visual Reliability $\times$ Angle $\times$ Repetition), totaling 80 trials. The unimodal T block followed an 8×5 factorial design (Angle $\times$ Repetition) for 40 trials. Finally, the bimodal VT blocks

followed a $2 \times 8 \times 3 \times 5$ factorial design incorporating visual reliability, angle, spatial subcondition, and repetitions.

Data analysis

Accuracy

Accuracy was defined as the spatial proximity of the participant's shot angle to the visual or tactile target. To account for the 360° spherical space, we first calculated the circular difference, $\Delta\theta$, between the shot angle (θ_{shot}) and the target angle (θ_{target}), constrained to the range $[-180, 180]^\circ$. Accuracy was then normalized to a $[0, 1]$ scale using the following transformation:

$$\text{Accuracy} = 1 - \frac{|\Delta\theta|}{180} \quad (5.1)$$

In this metric, an accuracy of 1.0 represents a perfect hit (zero angular error), while an accuracy of 0.5 represents a shot fired at a 90° offset. This normalized approach facilitates a direct comparison across different conditions and target directions.

Data were analyzed using Linear Mixed Models (LMM) to account for the nested structure of trials within participants. For all models, Participant ID was included as a random intercept to control for individual differences in baseline spatial performance.

To evaluate multisensory performance, we first conducted a 3-way LMM with Condition (Spatial Disparity: Congruent, Small, Large), Response Type (Visual, Tactile), and Visual Reliability (High, Low) as fixed effects. To investigate the potential impact of spatial region (Front, Side, Back) on performance, a second 4-way LMM was performed, incorporating Region as an additional fixed factor.

Directional Bias and Sensory Conflict

To quantify the degree to which the tactile distractor pulled the participant's response away from the visual target, we calculated the mean directional bias ($\bar{\theta}$) using circular statistics. For each condition, the angular errors were wrapped to the range $[-180, 180]^\circ$ and the circular mean was computed:

$$\bar{\theta} = \text{atan2} \left(\sum \sin(\Delta\theta), \sum \cos(\Delta\theta) \right) \quad (5.2)$$

A bias of 0° indicates a response centered on the target, while a bias shifting toward the distractor angle (45° or 135°) indicates multisensory capture. Additionally, we calculated the vector strength (R), also known as the resultant length, to measure response precision:

$$R = \frac{\sqrt{(\sum \sin \Delta\theta)^2 + (\sum \cos \Delta\theta)^2}}{N} \quad (5.3)$$

The R value ranges from 0 to 1, where 1.0 represents perfect directional consensus among trials and 0 represents a uniform random distribution. Significant differences in mean directional bias between reliability conditions were assessed using the Watson-Williams test, a circular analogue of a one-way ANOVA.

Movement Trajectories

Visualization To compare trajectories across conditions and subjects with varying reaction times (RT), movement data were standardized using a retrospective temporal alignment. For each trial, raw yaw values for the head and controller were unwrapped to a continuous angular scale and smoothed using a centered moving average filter (window = 20 frames; 60 Hz). Movement error was defined as the angular distance between the effector and the target, wrapped within $[-180^\circ, 180^\circ]$. To account for varying trial durations, trajectories were clipped from the start of trial (t_{start}) to the moment of the “shot” (t_{shot}). All trials were then interpolated onto a common 5-second temporal grid relative to the end of the trial (t_{shot}), allowing for subject-wise and group-level averaging.

Random Forest Classification To quantify the discriminability of movement patterns across conditions, we trained five independent Random Forest (RF) classifiers. Unlike linear models, the RF approach allowed us to identify non-linear interactions between kinematic features that characterize participant hesitation or compensation during multisensory processing. Feature sets were restricted to kinematics only; absolute angular error was excluded to ensure the models identified *how* participants moved rather than accuracy (*where* they ended). For each trial, we extracted four primary kinematic features for both the head and the controller (8 features total per input vector):

- **Max Velocity** (v_{max} , deg/s): The maximum rate of angular change, representing the ballistic intensity of the movement.

- **Peak Acceleration** (a_{peak} , deg/s^2): The maximum rate of velocity change; indicates the magnitude of the initial motor impulse.
- **Mean Jerk** ($\bar{j}$, deg/s^3): The average absolute rate of change of acceleration, serving as a proxy for movement smoothness and “shakiness”.
- **Path Length** (L , deg): The cumulative sum of absolute angular displacements, indicating the efficiency of the trajectory and the presence of corrective sub-movements.

All derivatives were computed using the unwrapped angular trajectories to avoid artifacts from circular wrap-around points.

The five classification tasks were defined as follows:

1. **Unimodal Baseline (V vs. T)**: Investigating whether visual and tactile-guided movements utilize distinct “motor languages” or control strategies.
2. **Intent Classification (Visual vs. Tactile response)**: Testing if top-down intent alters movement trajectories when participants are instructed to respond to a specific modality in bimodal conditions.
3. **Multisensory Conflict (Congruent vs. Small vs. Large)**: Testing for kinematic signatures (e.g., increased jerk) induced by spatial disparity between visual and tactile targets.
4. **Multisensory Enhancement (V vs. T vs. VT-congruent)**: Comparing unimodal performance against congruent bimodal performance to determine if the addition of a tactile belt and visual signal stabilizes the motor command.
5. **Visual Reliability (High vs. Low)**: Assessing how the degradation of visual signal quality changes the kinematic profile of the response.

Model Architecture and Evaluation Machine learning was implemented using a Random Forest classifier (100 estimators, maximum depth of 10). The dataset was partitioned into a 70/30 train-test split using stratified sampling. To address inherent imbalances in trial counts (e.g., specific reliability levels), we utilized balanced class weighting. This penalized misclassifications of the minority class more heavily, preventing bias toward more frequent trial types. Model performance was validated using a fixed random seed (42) to ensure reproducibility. To determine which

behavioral markers were most affected by multisensory variables, Gini Importance (Feature Importance) scores were calculated to rank which aspects of the kinematic features were most diagnostic for each classification task.

Model Fitting

We modeled multisensory integration using a Bayesian Causal Inference (BCI) framework to estimate the probability that visual (V) and tactile (T) cues originated from a common source ($C = 1$) or independent sources ($C = 2$). This framework allows us to derive forced-fusion, full-segregation, and maximum likelihood estimation (MLE) models as nested alternative hypotheses [29, 38, 88].

The BCI model assumes that a stimulus at a true spatial location s (or magnitude) results in noisy internal sensations x_V and x_T . These sensory observations are modeled as independent samples drawn from Gaussian distributions centered on the true stimulus positions s_V and s_T , with sensory noise defined by standard deviations σ_V and σ_T :

$$x_V \sim N(s_V, \sigma_V), x_T \sim N(s_T, \sigma_T) \quad (5.4)$$

Furthermore, the model incorporates a spatial prior representing the observer's expectations about stimulus locations. This prior is modeled as a Gaussian distribution centered at μ_P with standard deviation σ_P :

$$x_P \sim N(\mu_P, \sigma_P) \quad (5.5)$$

The observer must infer the underlying causal structure (C) that is hidden from the nervous system. The posterior probability of a common cause ($C = 1$) is calculated by combining the sensory evidence with a causal prior, P_{common} , according to Bayes' Rule:

$$p(C = 1|x_V, x_T) = \frac{p(x_V, x_T|C = 1)p_{common}}{p(x_V, x_T|C = 1)p_{common} + p(x_V, x_T|C = 2)(1 - p_{common})} \quad (5.6)$$

Here, P_{common} represents the observer's a priori belief that the visual and tactile signals belong to the same origin. Post-stimulus, the likelihood terms $p(x_V, x_T|C)$

quantify how likely the noisy internal observations x_V and x_T are, given each causal hypothesis.

If the signals are attributed to a common cause ($C = 1$), the optimal estimate of the stimulus location, $\hat{s}_{C=1}$, is a reliability-weighted average of both sensory signals and the spatial prior:

$$\hat{s}_{(V,C=1)} = \hat{s}_{(T,C=1)} = \hat{s}_{(VT,C=1)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_T}{\sigma_T^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_T^2} + \frac{1}{\sigma_P^2}} \quad (5.7)$$

If the signals are attributed to independent causes ($C = 2$), the estimates for each modality are computed separately. In this case, each estimate is influenced only by its corresponding sensory signal and the spatial prior:

$$\hat{s}_{(V,C=2)} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}, \hat{s}_{(T,C=2)} = \frac{\frac{x_T}{\sigma_T^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_T^2} + \frac{1}{\sigma_P^2}} \quad (5.8)$$

To obtain a final estimate of the number of visual or tactile stimuli, we assumed the observer combines the estimates from the two causal structures ($\hat{s}_{C=1}$ and $\hat{s}_{C=2}$) using one of three common decision strategies [88]:

Model Averaging (MA): The final estimate is a weighted sum of the estimates from both causal structures, where the weights are determined by their respective posterior probabilities:

$$\begin{aligned} \hat{s}_V &= p(C = 1|x_V, x_T)\hat{s}_{(V,C=1)} + p(C = 2|x_V, x_T)\hat{s}_{(V,C=2)} \\ \hat{s}_T &= p(C = 1|x_V, x_T)\hat{s}_{(T,C=1)} + p(C = 2|x_V, x_T)\hat{s}_{(T,C=2)} \end{aligned} \quad (5.9)$$

Model Selection (MS): The observer selects the estimate from the causal structure with the higher posterior probability. If $p(C = 1|x_V, x_T) > 0.5$, the observer adopts $\hat{s}_{C=1}$; otherwise, they adopt $\hat{s}_{C=2}$:

$$\begin{aligned} \hat{s}_V &= \begin{cases} \hat{s}_{V,C=1}, & \text{if } p(C = 1|x_V, x_T) > 0.5 \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_T &= \begin{cases} \hat{s}_{T,C=1}, & \text{if } p(C = 1|x_V, x_T) > 0.5 \\ \hat{s}_{T,C=2}, & \text{otherwise} \end{cases} \end{aligned} \quad (5.10)$$

Probability Matching (PM): A stochastic strategy where the observer chooses a structure proportional to its posterior probability. On each trial, a threshold ξ is drawn from a uniform distribution $\mathcal{U}(0, 1)$. If $p(C = 1|x_V, x_T) > \xi$, the observer chooses $\hat{s}_{C=1}$:

$$\begin{aligned}\hat{s}_V &= \begin{cases} \hat{s}_{V,C=1}, & \text{if } p(C = 1|x_V, x_T) > \xi \\ \hat{s}_{V,C=2}, & \text{otherwise} \end{cases} \\ \hat{s}_T &= \begin{cases} \hat{s}_{T,C=1}, & \text{if } p(C = 1|x_V, x_T) > \xi \\ \hat{s}_{T,C=2}, & \text{otherwise} \end{cases}\end{aligned}\tag{5.11}$$

We evaluated the BCI framework against three alternative models, treating them as restricted or special cases of the full BCI model: (i) Forced Fusion: Assumes a common cause is always present ($P_{common} = 1$), (ii) Full-Segregation: Assumes signals are always independent ($P_{common} = 0$), (iii) Maximum Likelihood Estimation (MLE): A special case of forced-fusion that assumes no prior spatial/numerosity information (effectively $P_{common} = 1$ and a non-informative flat prior, $\sigma_P \rightarrow \infty$). An overview of these models is presented in Table 5.1.

To account for the circular nature of the 360° environment while using a Gaussian BCI framework, we transformed all raw shot angles into a relative coordinate system. All visual stimulus locations (s_V) were normalized to 0°, with tactile stimulus locations (s_T) and participant responses ($\text{resp}_V, \text{resp}_T$) expressed as angular offsets ranging from -180° to $+180^\circ$. In this reference frame, the tactile stimulus s_T was positioned at 0° , $\pm 45^\circ$, or $\pm 135^\circ$ relative to the visual center.

To investigate multisensory integration under conditions where participants relied primarily on peripheral vision rather than overt head turning, BCI model fitting was restricted to the Front region of the 360° space. This region provided the most reliable visual performance and best reflected the intended manipulation of visual reliability, whereas targets in the Side and Back regions could often be resolved through exploratory head movements rather than immediate multisensory inference. Accordingly, model fitting was performed using only trials from the Front region to estimate participant-specific parameters ($\sigma_V, \sigma_T, \sigma_P, P_{common}$).

We performed parameter estimation using the BCI Toolbox [89]. For each fit, we utilized the Bayesian Information Criterion (BIC) to identify the best-fitting decision strategy (Model Averaging, Selection, or Probability Matching) and to compare the full BCI model against the restricted alternative models (Forced-Fusion,

Full-Segregation, and MLE). To evaluate model goodness-of-fit, we calculated McFadden’s pseudo- R^2 (R_{MF}^2) measure based on the log-likelihood of the observed data [43]. For each subject and condition, we defined a null model by pooling data across all experimental conditions to generate a marginal probability distribution for each response channel. The null log-likelihood (LL_{null}) was then compared to the log-likelihood of each candidate model (LL_{model}). The final R_{MF}^2 was defined as:

$$R_{MF}^2 = 1 - \frac{LL_{model}}{LL_{null}} \quad (5.12)$$

where a value of 0.2-0.4 indicates very good model fit.

Table 5.1: Overview of models and their free parameters

Model	Free Parameters	#
Bayesian Causal Inference (BCI)	$p_{common}, \mu_P, \sigma_P, \sigma_V, \sigma_T$	5
Forced-fusion	$\mu_P, \sigma_P, \sigma_V, \sigma_T$	4
Full-segregation	$\mu_P, \sigma_P, \sigma_V, \sigma_T$	4
Maximum-likelihood estimation (MLE)	σ_V, σ_T	2

5.3 Results

Accuracy

Unimodal Baselines To establish the baseline reliability of each sensory modality, a Linear Mixed Model (LMM) was conducted on the mean accuracy of unimodal conditions. The analysis revealed a significant main effect of visual reliability condition ($\chi^2(2) = 373.12, p < .001$), with the high visual reliability condition yielding the highest accuracy (0.96 ± 0.08), significantly outperforming the tactile-only (no visual) condition (0.90 ± 0.07 ; $z = 4.65, p < 0.001$; Fig. 5.5 top row). Conversely, the low visual reliability condition resulted in significantly lower accuracy than the tactile-only condition ($0.74 \pm 0.19, z = -13.91, p < 0.001$).

Visual-tactile Integration under Spatial Disparity Next, to examine how varying spatial discrepancies and visual reliability affect performance, a 3-way LMM was conducted on accuracy scores within the multisensory (VT) conditions (Fig. 5.5 rows 2-4). The analysis revealed significant main effects for Condition ($\chi^2(2) = 21.88, p < .001$) and Response Type ($\chi^2(1) = 28.47, p < .001$). Crucially, a significant 3-way interaction was observed between Condition, Response Type, and Reliability ($\chi^2(2) = 153.46, p < .001$; Fig. C.3).

In the VT-congruent condition, accuracy remained high across all visual reliability levels and response modalities (Visual response: *High* 0.99 ± 0.01 , *Low* 0.94 ± 0.08 ,

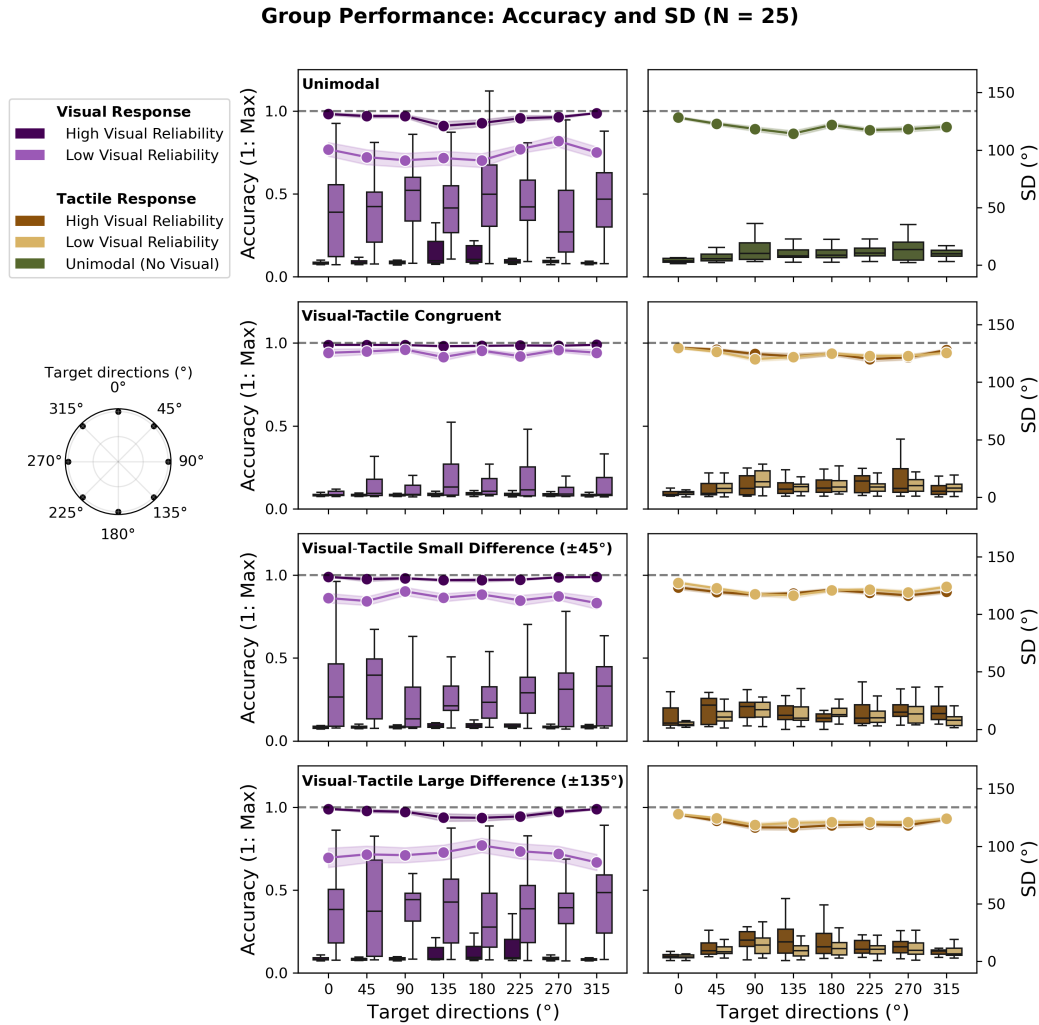


Figure 5.5: Group Performance ($N = 25$): Accuracy and Standard Deviation (SD) Across Visual Reliability Conditions. Line plots represent mean accuracy ($[0, 1]$) across conditions; box plots illustrate the distribution of response standard deviation (SD). For visual responses (participants shoot at UFO), dark purple denotes high visual reliability trials, while light purple denotes low visual reliability. For tactile responses (participants shoot at direction of vibration), brown and yellow represent high and low visual reliability, respectively. Green indicates unimodal tactile control trials (no visual stimuli). See Fig. C.2 for mean accuracy pooled across all directions.

tactile response: *High* 0.94 ± 0.06 , *Low* 0.93 ± 0.05). However, the impact of visual reliability became evident as spatial disparity increased. In particular, for visual responses in low reliability conditions, accuracy dropped to 0.86 ± 0.13 for VT-small and 0.72 ± 0.24 for VT-large. This indicates that when visual reliability is low, large spatial conflicts with tactile cues severely degrade visual localization. Conversely, when visual reliability was high, participants maintained high accuracy

(0.96 ± 0.07) even in the presence of a large spatial discrepancy, suggesting the visual signal dominated the percept and ignored the conflicting tactile information.

Interestingly, tactile responses showed a much higher degree of resilience to visual reliability manipulations. In both VT-small and VT-large conditions, tactile accuracy remained remarkably stable (ranging from 0.90 to 0.92) regardless of visual reliability levels. The significant 2-way interaction between Response Type and Reliability ($\chi^2(1) = 8.98, p = .003$) confirms that the performance drop associated with low visual reliability was significantly more pronounced for visual responses than for tactile responses.

Effects of Spatial Region To determine if sensory reliability interacted with the physical location of the target, we analyzed performance across three regions: Front, Side, Back (Fig. 5.3A). A LMM on unimodal trials revealed a significant main effect of Condition ($\chi^2(2) = 171.13, p < .001$) and marginal significance for Region ($\chi^2(2) = 5.70, p = 0.058$). Participants were most accurate when targets were located in the Front region. Specifically, the tactile baseline accuracy in the front (0.93 ± 0.06) was significantly higher than accuracy in the Back ($0.89 \pm 0.07; z = -2.20, p = .028$). Interestingly, there is no significant interaction between Region and Condition ($\chi^2(4) = 3.85, p = 0.43$). This suggests that while there is a general “penalty” for targets located behind or to the side of the participant, the relative hierarchy of the sensory reliability (Visual High > Tactile N/A > Visual Low) remains consistent regardless of direction.

For multisensory (VT) trials, a 4-way LMM was conducted to evaluate the influence of Region alongside Condition, Response Type, and Visual Reliability. Same as before, we observed a significant main effect of Condition ($\chi^2(2) = 14.36, p < .001$). Unlike the unimodal baseline, multisensory trials showed a significant main effect of Region ($\chi^2(2) = 10.45, p = .005$). Targets located in the Front yielded the highest overall accuracy ($M = 0.96$ at the intercept). Performance significantly declined as targets moved to the Side ($z = -0.040, p = .015$) and reached its lowest point in the Back region ($z = -0.044, p = .003$). This $\sim 4\%$ drop in accuracy suggests that multisensory integration is slightly less efficient when the body must orient toward targets outside the primary field of view.

The analysis also revealed a highly significant 3-way interaction between Condition, Response Type, and Reliability ($\chi^2(2) = 82.34, p < .001$). This confirms that the degree of shooting accuracy depends on both the spatial disparity and the reliability

of the visual signal. Crucially, the 4-way interaction (Condition, Response Type, Region, and Visual Reliability) was not significant ($p = .063$), though it trended toward significance (Fig. C.4). This suggests that while overall accuracy is lower in the Side and Back, the fundamental way participants integrate visual and tactile cues is spatially invariant within the VR environment.

Directional Bias and Sensory Conflict

We observed a significant interaction between spatial disparity and visual reliability on directional bias (Fig. 5.6). When visual reliability was high, participants exhibited near-perfect target localization with negligible bias ($\sim 0^\circ$) and high precision ($R > 0.95$). However, under low visual reliability, responses were significantly pulled by the tactile distractor. When the difference is small ($\pm 45^\circ$), responses were biased by 7.73° toward the tactile target. When the difference is large ($\pm 135^\circ$), the bias increased to 11.24° , though this was accompanied by a significant decrease in response precision ($R = 0.44$). This suggests that participants did not simply re-weight their internal estimate toward the tactile target. Instead, large spatial disparity and high visual noise lead to confusion and sub-optimal causal inference, thus the significant increase in response variance.

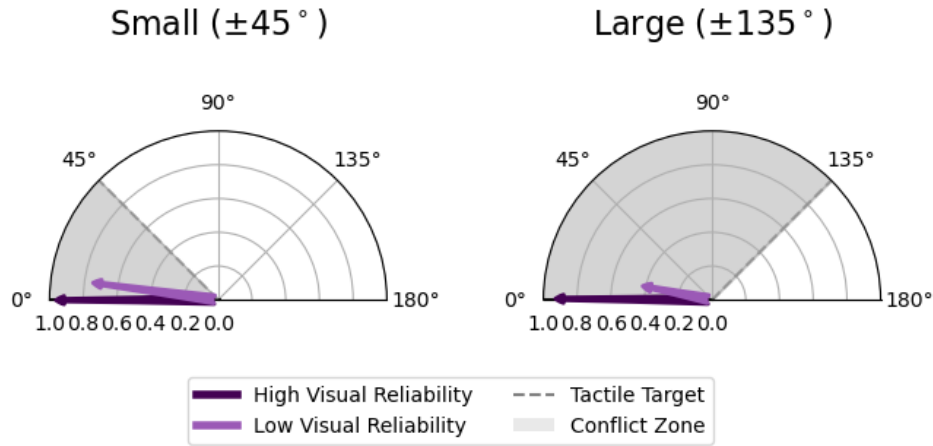


Figure 5.6: **Mean Directional Bias and Response Precision across Spatial Disparity and Visual Reliability** Polar plots illustrate the aggregate response vectors for visual responses. The angular position of each arrow represents the mean directional bias ($\bar{\theta}$), where 0° (horizontal) denotes the visual target and the gray shaded region represents the conflict zone toward the tactile target. The length of each arrow represents the vector strength (R), a measure of circular precision ranging from 0 (random) to 1 (perfect consensus).

Movement Trajectories

Group-averaged normalized movement trajectories demonstrate that across all conditions, subjects successfully converged on the target (0°) prior to the shot. The most prominent effect of condition and visual reliability was observed in the temporal domain rather than the spatial path (Fig. 5.7).

When participants targeted tactile stimuli (Brown/Yellow/Green dashed lines), mean reaction times (RT) were consistently lower (~ 1.17 s) and relatively invariant to visual reliability levels. Conversely, targeting visual stimuli under low reliability (Light Purple dashed lines) resulted in significantly longer search times, with the unimodal visual condition yielding the highest mean RT (3.83 s). Across the multisensory conditions, a clear hierarchy emerged: VT-congruent trials exhibited the shortest latencies, followed by VT-small and VT-large, with unimodal trials requiring the longest searching time.

Random Forest Classification

Because group-averaged trajectories exhibited only subtle spatial divergence, we reasoned that the underlying control strategies were better captured through discrete kinematic signatures rather than raw positional data. We extracted four kinematic features, including: max velocity (v_{max} , deg/s), peak acceleration (a_{peak} , deg/s^2), mean jerk ($\bar{j}$, deg/s^3), and path length (L , deg). These features served as inputs for Random Forest classifiers trained to decode five distinct experimental dimensions: unimodal baseline (V vs. T), task intent (Visual vs. Tactile response), spatial discrepancy (Congruent vs. Small vs. Large), multisensory enhancement (V vs. T vs. VT-congruent), and visual reliability levels. This multivariate approach allows for the detection of non-linear kinematic patterns that distinguish sensory processing strategies that are otherwise obscured in averaged spatial plots. The performance metrics for these models are summarized in Table 5.2.

High Decodability in Modality and Intent The classification results revealed a clear hierarchy in the decodability of movement kinematics. The model achieved high accuracy in distinguishing between unimodal baseline conditions (V vs. T; 88%, $F1 = 0.86$) and in decoding task intent (85%, $F1 = 0.85$). Confusion matrices for these models exhibited strong and balanced performance; for example, the unimodal baseline model accurately predicted true visual trials 90% of the time and true tactile trials 83% of the time (Fig. 5.8, top row). This suggests that the motor

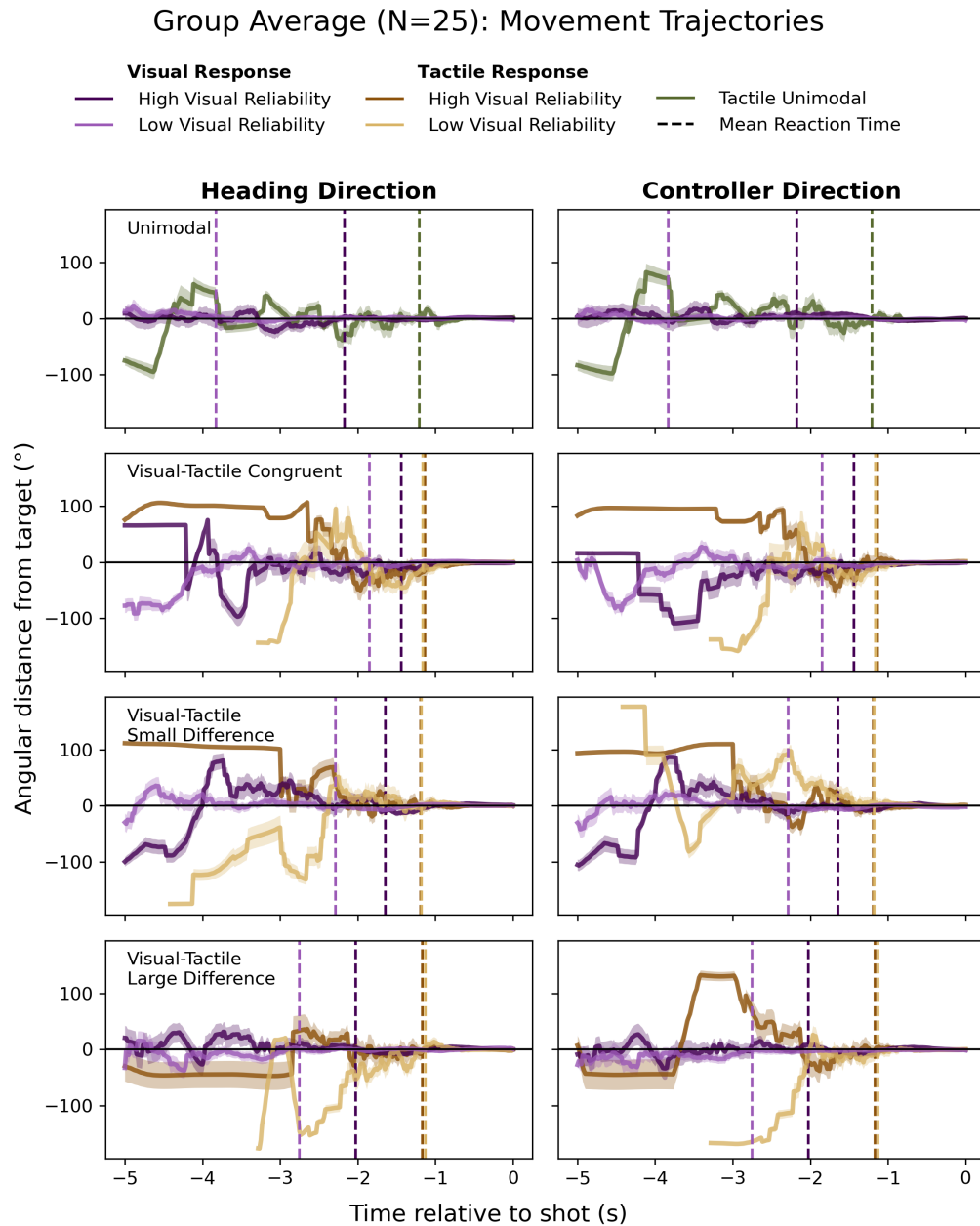


Figure 5.7: Group Average Movement Trajectories across Conditions and Visual Reliability Levels. Angular distance from the target ($^{\circ}$) is plotted relative to shot time (0s) for both Heading Direction (Left) and Controller Direction (Right). Solid lines indicate the group mean ($N = 25$), with shaded regions representing ± 1 SEM. Vertical dashed lines indicate the mean reaction time for each respective condition and reliability level. Negative time values represent the approach phase preceding the final shot event.

Table 5.2: **Performance Metrics for Random Forest Kinematic Classifiers.** Classification accuracy and F1 scores are reported for each experimental dimension.

Classification Task	Classes	Accuracy (%)	F1 Score ^a
Unimodal Baseline	V/T	88	0.86
Task Intent	V/T	85	0.85
Spatial Discrepancy	Congruent/Small/Large	40	0.40
Multisensory Enhancement	V/T/VT-Congruent	66	0.63
Visual Reliability	High/Low	59	0.59

^a The F1 score is the harmonic mean of precision and recall, given by $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$, where $\text{Precision} = \frac{TP}{TP+FP}$ and $\text{Recall} = \frac{TP}{TP+FN}$. *TP*: True Positive; *FP*: False Positive; *FN*: False Negative.

system employs distinct, identifiable control strategies based on the primary sensory source.

Robustness to Spatial Discrepancy and Visual Reliability Classifiers tasked with decoding finer sensory nuances, such as spatial discrepancy and visual reliability, exhibited lower performance. The spatial discrepancy classifier achieved only 40% accuracy (chance = 33.3%), indicating that while reaction times may vary, the physical execution of the movement remains relatively robust to the size of the discrepancy (Fig. 5.8, third row). Similarly, visual reliability was decoded at 59% (chance = 50%), suggesting that while participants reported increased difficulty under low-reliability conditions (longer reaction times), the resulting motor output maintained a high degree of kinematic consistency with high-reliability trials (Fig. 5.8, bottom row).

Evidence of Tactile Dominance Interestingly, the classification of multisensory enhancement revealed a notable asymmetry indicating a strong tactile dominance. While the model successfully separated unimodal tactile trials from unimodal visual trials, unimodal tactile trials were frequently misclassified (86% of the time) as VT-congruent trials (Fig. 5.8, fourth row). This high degree of overlap implies that participants' movement trajectories in a multisensory environment are heavily dominated by tactile signals, and visual information exerts a negligible influence on the physical execution of the movement.

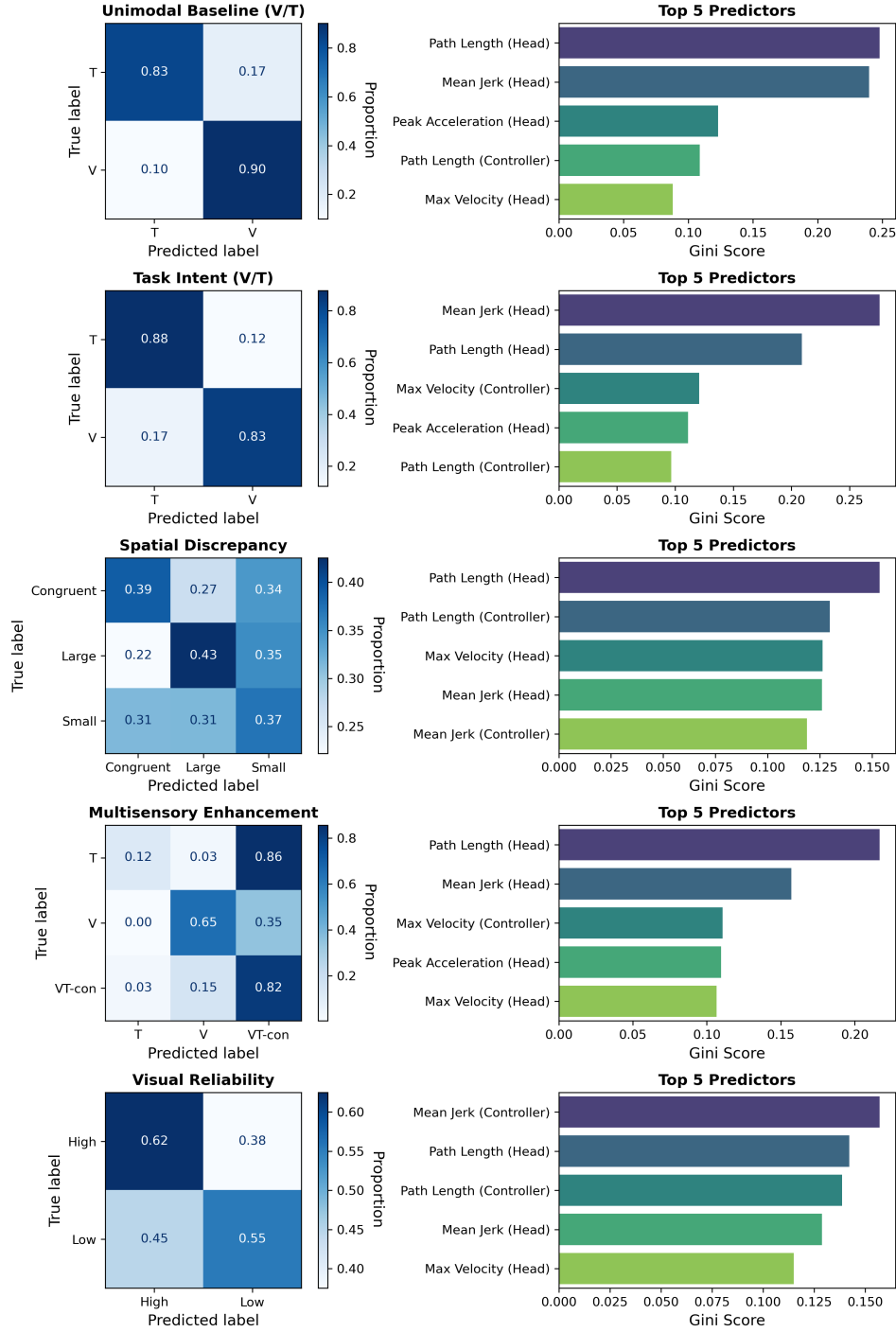


Figure 5.8: Random Forest Classification Performance and Kinematic Feature Importance. Each row represents a distinct classification task performed on movement kinematics ($N = 25$). **Left column:** Normalized confusion matrices illustrating the proportion of trials classified for each class label; dark diagonals indicate high classification accuracy. **Right column:** Top 5 predictors for each model ranked by Gini Importance, which quantifies a feature's contribution to the model. Feature labels indicate whether the data was recorded from the *Head* or *Controller* sensor.

Kinematic Feature Importance An analysis of Gini importance across all five models revealed that Path Length (Head) and Mean Jerk (Head) were consistently the most discriminative predictors of sensory and task-related states. In the high-accuracy models (Unimodal Baseline and Task Intent), these two features contributed nearly 50% of the total cumulative importance. This suggests that the spatial efficiency (path length) and smoothness (mean jerk) of the head trajectory serve as the primary kinematic markers of sensory modality. Notably, while head kinematics dominated baseline and intent decoding, smoothness of the controller trajectory (mean jerk) emerged as the top predictor for classifying visual reliability.

Model Fitting

To quantify how participants integrated visual and tactile cues, we applied a Bayesian Causal Inference (BCI) framework. We compared the full BCI model against three restricted models: Forced-Fusion (FF), Full-Segregation (FS), and Maximum Likelihood Estimation (MLE) [29, 38, 88].

Model performance was evaluated using the Bayesian Information Criterion (BIC) and McFadden’s Pseudo- R^2 [43]. To validate these metrics, we performed a correlation analysis between individual regional model fits and observed behavioral accuracy.

Model Comparison

Analysis of BIC scores on model fits in Front region revealed a reliability-dependent shift in integration strategies. Under high visual reliability, the full BCI model emerged as the winning model, demonstrating the lowest mean BIC (141 ± 4) and significantly outperforming FS ($p_{adj} = 0.015$), MLE ($p_{adj} < 0.001$), and FF ($p_{adj} < 0.001$) models (Fig. 5.9 Left; Friedman Test: $\chi^2 = 68.09$, $p < 0.001$, post hoc pairwise comparisons with Bonferroni correction). The BCI model also captured the most variance, with a mean McFadden Pseudo- R^2 (R_{MF}^2) of 0.36 ± 0.02 (0.2-0.4 indicates very good model fit).

Under low visual reliability, the simpler MLE model (2 free parameters) achieved a lower mean BIC (272 ± 8) than the full BCI model (290 ± 12), but this difference is not statistically significant ($p_{adj} = 0.256$). However, this numerical advantage was not reflected in the variance explained. While the BCI model maintained a robust R_{MF}^2 of 0.22 ± 0.01 , the MLE model’s R_{MF}^2 collapsed to near-zero (0.02 ± 0.02). This suggests that while the MLE model is mathematically more parsimonious under low

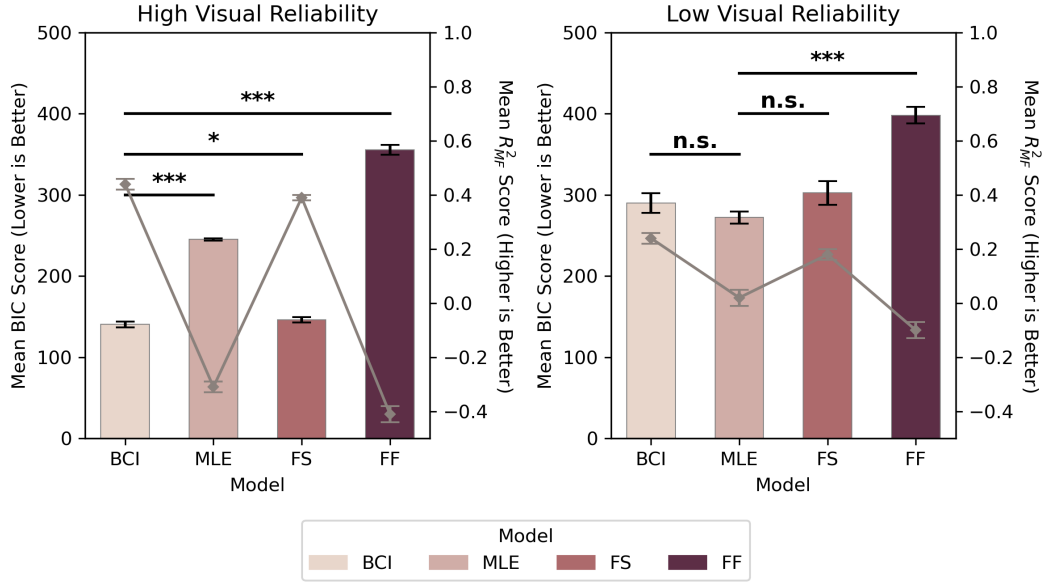


Figure 5.9: **Model Comparison by Visual Reliability.** Mean Bayesian Information Criterion (BIC) scores (bar plots, left axis; lower is better) and mean R^2_{MF} (line plots, right axis; higher is better) for the four models tested: BCI, MLE, FS, FF. Error bars represent ± 1 standard error of the mean (SEM). Asterisks denote statistical significance (n.s. Not Significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

visual reliability, it fails to capture the underlying structure of the behavioral data. An alternative explanation is that behavior under high uncertainty is inherently noisy (See Fig. 5.6 red arrows), and response precision was significantly reduced under low visual reliability. Because participants did not behave in any statistically optimal fashion (as the four models would assume), model fits were generally suboptimal; since BIC penalizes models with more free parameters, the BCI model receives a heavier penalty.

Predictive Validity of BCI and MLE models To adjudicate between the two competing models (BCI and MLE), we examined the relationship between a model's global fit and participants' mean unimodal accuracy (Fig. 5.10). We don't observe significant trends under high visual reliability, as unimodal accuracy was near ceiling. Data points were clustered around 1.0. However, for visual responses under low visual reliability and tactile responses, we found that as behavioral accuracy increases, model fit improves. This is evident in the significant negative correlations between BIC scores and accuracy for the Bayesian Causal Inference Model (Visual response: $r = -0.68$, $p < 0.001$; Tactile response: $r = -0.63$, $p < 0.001$). This trend is mirrored by the goodness-of-fit measure (R^2_{MF}). We observe significant

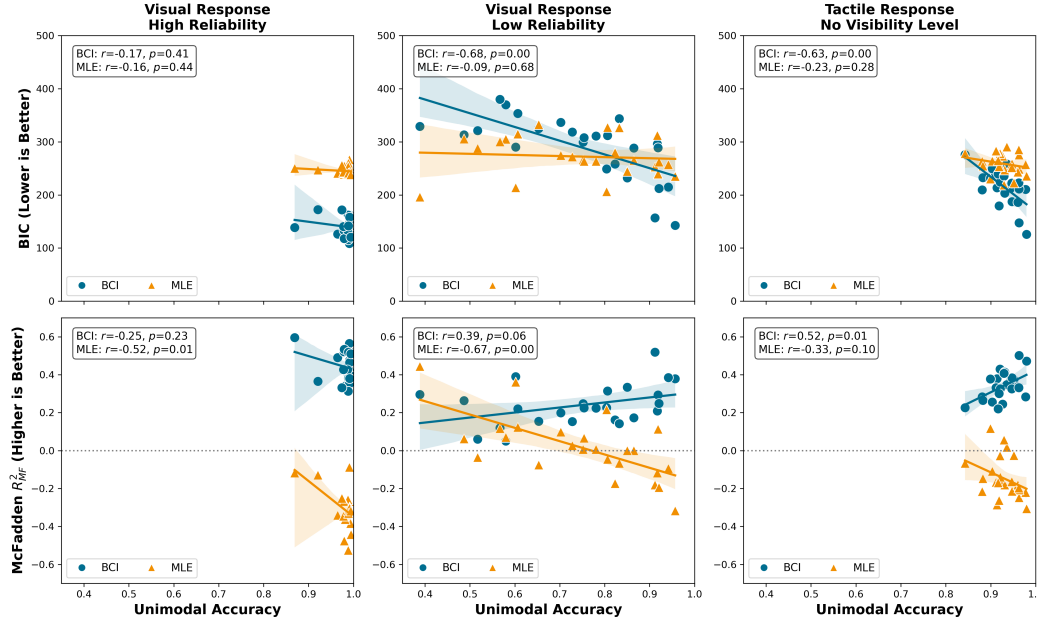


Figure 5.10: **Correlation between behavioral accuracy and BIC scores (Model performance; the lower the better), and R^2_{MF} values (Goodness of fit; the higher the better).** Each data point shows a participant’s mean visual/tactile accuracy versus their BIC score and R^2_{MF} values from BCI and MLE model fitting. Blue trend lines represent BCI model, orange trend lines represent MLE model.

positive correlations between R^2_{MF} and accuracy for tactile response ($r = 0.52$, $p = 0.01$). We observe a positive trend under low visual reliability as well, but it is marginally significant ($r = 0.39$, $p = 0.06$).

In stark contrast, the MLE model showed no such predictive validity. Despite its numerically lower BIC scores in the low-reliability visual condition, those scores showed no significant correlation with observed accuracy ($r = -0.09$, $p = 0.68$). Furthermore, MLE R^2_{MF} values were consistently lower than those of the BCI model. Most notably, for visual response under high visual reliability and tactile response, the MLE model yielded negative R^2_{MF} scores, indicating a fit worse than that of the null model. This disconnect identifies the MLE “win” as a mathematical artifact of the BIC complexity penalty; the model appears parsimonious simply by having fewer parameters, but it fails to track the systematic cognitive processes driving multisensory performance.

Parameter Estimation

Having established that the BCI model provides the most psychologically plausible and predictive account of multisensory behavior, we analyzed the best-fitting

parameters to characterize the underlying computational processes.

Manipulating visual reliability had a massive and expected impact on visual sensory noise (Table 5.3, Fig. C.5). Visual noise (σ_V) increased dramatically from high reliability condition to low reliability condition ($t(24) = -8.09$, $p_{adj} < .001$, $BF_{10} = 7,528,019$). Accompanying this degradation in visual sensitivity, the width of the spatial prior (σ_P) also significantly broadened under low reliability compared to high reliability ($t(24) = -4.07$, $p_{adj} < .01$, $BF_{10} = 22.57$). The expansion of the prior width (σ_P) under low visual reliability replicates the adaptive recalibration observed in **Chapter 4**. This suggests that the brain widens its spatial expectations whenever visual evidence becomes unreliable.

Table 5.3: Fitted BCI Parameters for Front region. Mean parameter values ± 1 standard error of the mean (SEM) across high and low visual reliability conditions. Angular parameters (σ_V , σ_T , σ_P , μ_P) are expressed in degrees.

Region	Parameter	High Reliability (Mean $\pm$ SE)	Low Reliability (Mean $\pm$ SE)	t	p_{adj}	BF_{10}
Front	p_{common}	0.42 ± 0.05	0.42 ± 0.05	0.04	1.00	0.21
	σ_P ($^\circ$)	60.68 ± 4.12	84.49 ± 5.03	-3.55	<.01	22.57
	μ_P ($^\circ$)	48.77 ± 1.18	22.86 ± 4.00	6.48	<.001	1.6e4
	σ_V ($^\circ$)	3.14 ± 0.38	73.68 ± 7.36	-9.43	<.001	7.5e6
	σ_T ($^\circ$)	27.50 ± 2.54	17.69 ± 3.14	2.35	0.14	2.07

Note: p_{adj} reflects Bonferroni correction for 5 comparisons. $BF_{10} < 1$ indicates evidence favoring the null hypothesis.

Importantly, as illustrated by the response distribution curves (Fig. C.6), these spatial priors are extremely flat across the 360° space relative to the sensory likelihoods. Because an uninformative (flat) prior exerts minimal pull on the final posterior estimates, comparing the center of these distributions (μ_P) holds very little behavioral significance or interpretive value. See Fig. 5.11 for group average BCI model fits, and Figs. C.7, C.8, C.9 for group average MLE, FS, and FF model fits.

Interestingly, participants' underlying bias to infer a common cause (p_{common}) did not differ between high (0.42 ± 0.05) and low (0.42 ± 0.05) visual reliability conditions globally ($t(24) = 0.035$, $p_{adj} = 1.0$, $BF_{10} = 0.21$). Tactile noise (σ_T) showed a slight drop under low visual reliability, but the difference is not statistically significant ($t(24) = 2.35$, $p_{adj} = 0.14$, $BF_{10} = 2.07$). Furthermore, an analysis of the BCI decision strategy selection distribution indicated that "Model Selection" was the most frequently adopted strategy globally under high visual reliability (53%),

followed by “Model Averaging” (36%). Under low visual reliability, “Model Averaging” became the most prominent (40%), followed closely by “Model Selection” (35%) and “Probability Matching” (25%).

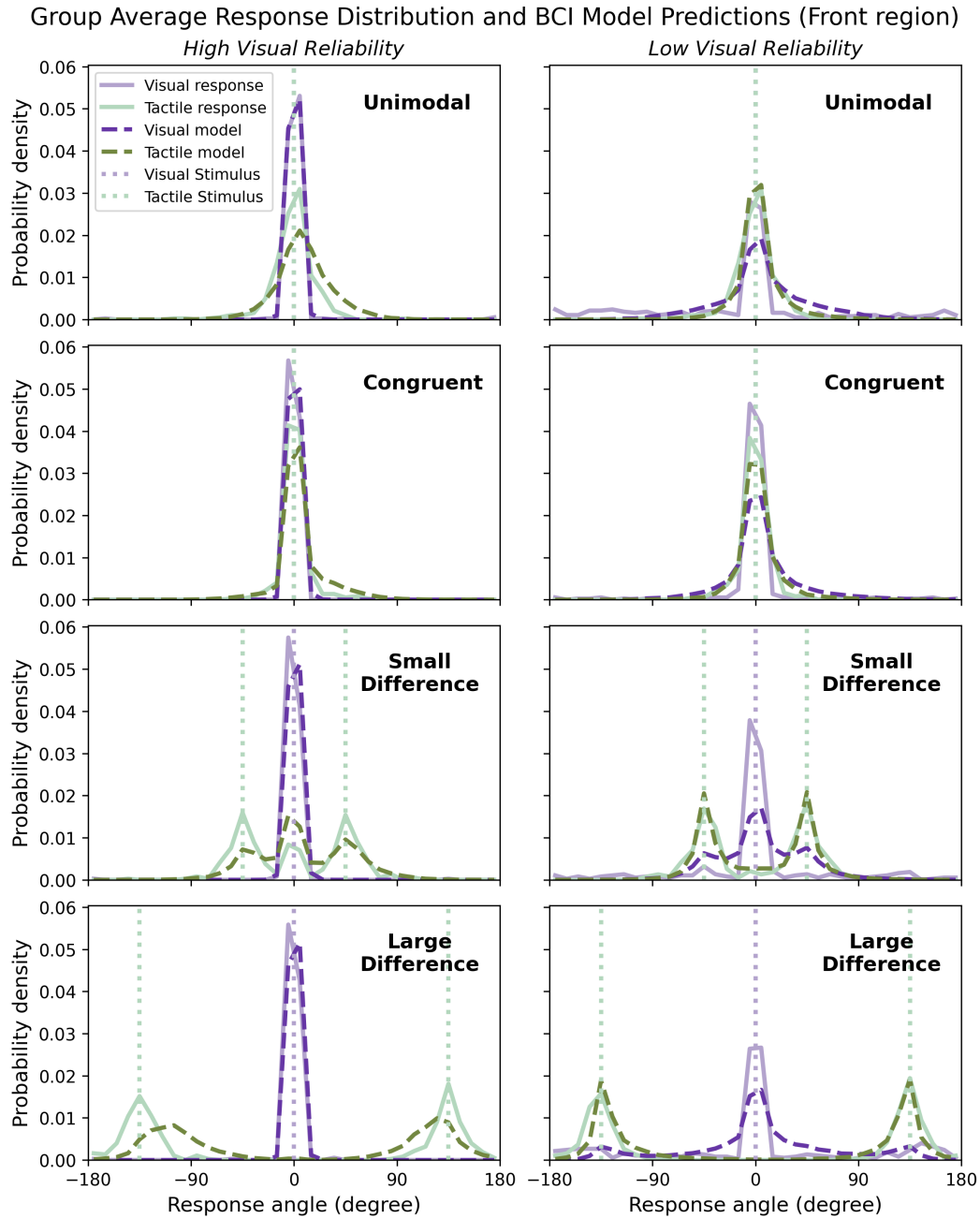


Figure 5.11: **Group average Bayesian Causal Inference model fits for Front region.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° .

Model Identifiability and Parameter Recovery To ensure the reliability of these parameter estimates, we performed a comprehensive validation of the model’s recovery performance and identifiability (see Appendix B, Figs. C.10, C.11).

First, parameter recovery analyses using synthetic datasets demonstrate robust recovery for key parameters of interest (σ_V, σ_T ; Fig. C.10). Recovery correlations for these primary values were consistently high (σ_V : $R^2 = 0.91$, σ_T : $R^2 = 0.63$), indicating that the model successfully captures the generative process. While p_{common} displays a lower R^2 (0.43), its remarkably low Mean Squared Error ($MSE = 0.04$) indicates that estimates are highly accurate but clustered within a narrow range, making R^2 sensitive to minor deviations. The relatively poor recovery of spatial prior parameters (σ_P, μ_P) was expected because the fitted priors were sufficiently broad to be uninformative, the model’s likelihood surface is relatively flat in these dimensions, rendering it less sensitive to their specific values without compromising the estimation of other parameters.

Second, we examined the inter-parameter correlation structure to determine if the model could uniquely isolate sensory and prior components. Across both global and regional analyses, the parameters exhibited minimal inter-parameter correlations (Fig. C.11), confirming high model identifiability. While specific correlations were observed between sensory noise (σ_V, σ_T) and the spatial prior mean (μ_P), these are considered functionally negligible. As established, the uninformatively broad spatial prior (σ_P) renders these specific mean offsets behaviorally inert. Collectively, these validation steps confirm that our experimental design provides sufficient data density to uniquely identify the key parameters driving multisensory integration. This provides high confidence in our central finding: that the underlying causal architecture remains structurally stable while sensory and spatial uncertainties adapt to local reliability.

5.4 Discussion

The present study investigated the computational mechanisms and continuous motor strategies underlying visual-tactile integration across a fully immersive 360° space. By systematically manipulating spatial discrepancy and visual reliability, we demonstrate that the explanatory power of Bayesian Causal Inference (BCI) is strongly dependent on visual reliability. While formal model comparison occasionally favored a simpler Maximum Likelihood Estimation (MLE) fusion model under extreme visual degradation due to its parsimony, this did not reflect a better

account of behavior. Instead, BCI remained the only model whose fit systematically tracked behavioral performance. BCI modeling revealed that while tactile precision varied across the body, the fundamental binding tendency remained spatially invariant across the 360° environment. In addition, kinematic analyses revealed a dissociation between perception and action, with early motor behavior dominated by tactile signals despite visually dominated perceptual endpoints.

Visual Reliability Dictates Computational Strategy When visual reliability was high, participants' behavior was well captured by the full BCI model ($R_{MF}^2 = 0.36 \pm 0.02$). Under low visual reliability, however, model performance degraded substantially. Although MLE yielded lower BIC values in these high-noise conditions, it exhibited a floor effect in explained variance ($R_{MF}^2 \approx 0.02$) and in some cases negative McFadden pseudo- R^2 , indicating performance worse than a null model. This dissociation highlights a critical distinction between relative model comparison and absolute explanatory power.

Crucially, BCI fits scaled with behavioral accuracy. BCI BIC scores were strongly correlated with unimodal behavioral accuracy in low reliability conditions ($r = -0.68$), whereas MLE showed no such relationship ($r = -0.09$). Thus, even when overall fit degrades, BCI continues to capture systematic variation in behavior, while MLE does not. These results suggest that BCI is not replaced under high uncertainty; rather, its effective descriptive power depends on the reliability of behavior itself.

This scaling relationship provides important insight into the question of compensation under degraded vision. If observers compensated for reduced visual reliability by increasing their tendency to bind cues, one would expect systematic changes in the binding tendency parameter (p_{common}). However, we observed that p_{common} remained relatively stable across reliability conditions (0.42 vs. 0.42). This suggests that the brain does not compensate for degraded vision by increasing cross-modal binding. Instead, the causal inference structure remains stable, while uncertainty is absorbed through changes in sensory likelihoods and priors [10, 48, 90].

Together, these findings indicate that under extreme uncertainty, behavior deviates from the assumptions required for stable recovery of Bayesian computations. This is consistent with the distinction between intrinsic and extrinsic uncertainty [6]: while the nervous system can account for intrinsic noise, externally imposed uncertainty (such as experimental manipulation in this study) may be difficult to represent or learn. As a result, responses become more variable and less consistent with ideal

observer predictions, leading to degraded model fits despite the persistence of an underlying inferential architecture.

Tactile Dominance in Motor Execution While the BCI model captured perceptual endpoints, kinematic analyses revealed a dissociation between perception and action. Under high visual reliability, visual uncertainty ($\sigma_V \approx 4^\circ$) was substantially lower than tactile uncertainty ($\sigma_T \approx 38^\circ$), predicting visual dominance in perceptual estimates. However, early motor trajectories exhibited strong tactile dominance, with unimodal tactile trials frequently misclassified as congruent bimodal trials (86%).

This pattern suggests a modular, two-stage process in 360° spatial behavior. Tactile signals, encoded in egocentric coordinates, provide rapid orienting information for initial motor alignment, while visual input (when available) supports later-stage refinement of perceptual estimates [52, 75, 86]. Rather than reflecting a single integrated estimate, behavior appears to arise from sequential or parallel computations operating on different time scales and reference frames.

Implications and Future Directions These findings have important implications for both theory and application. Theoretically, they suggest that multisensory integration is not characterized by dynamic compensation via increased binding under degraded vision. Instead, the brain maintains a stable causal inference structure, while the quality of sensory representations determines how effectively this structure can be expressed in behavior. This reframes BCI as a reliability-dependent framework, rather than a universally optimal model.

From an applied perspective, the dissociation between tactile-driven motor behavior and visually refined perception suggests that sensory substitution systems can leverage low-resolution tactile cues to guide action effectively, without requiring high-fidelity spatial encoding [12, 39, 73].

Finally, the reduced model fits observed under high uncertainty point to limitations of current modeling approaches. Incorporating circular statistics (e.g., Von Mises distributions) and lapse parameters may better capture the variability observed in 360° tasks. More broadly, identifying the mechanisms underlying behavior in this high-uncertainty regime remains an open challenge, highlighting the need for theoretical extensions beyond current formulations of multisensory inference.

5.5 Conclusion

Our results demonstrate that the Bayesian Causal Inference (BCI) framework provides the most consistent and robust account of multisensory behavior across varying levels of visual degradation. Although extreme uncertainty can numerically favor simpler models such as maximum likelihood estimation (MLE) due to their lower complexity, BCI uniquely captures the underlying computational structure of perception. Importantly, the stability of the p_{common} parameter across reliability levels and spatial regions suggests that the prior probability of sensory binding is an inherent property of the observer, largely independent of moment-to-moment changes in sensory noise.

Critically, BCI model fit scales with task performance: observers whose behavior more closely approximates Bayes-optimal inference achieve higher accuracy, reflecting more effective integration of sensory information. This relationship indicates that reduced model fit under degraded visual conditions does not reflect a breakdown of the inference process itself, but rather a limitation in how well sensory reliability can be estimated and incorporated into the model.

Kinematic analyses further support this interpretation by revealing a dissociation between the inferred percept and its behavioral implementation. While BCI captures the structure of perceptual inference, continuous movement trajectories are systematically biased toward tactile signals, particularly during early stages of action. This suggests that multisensory integration operates within a common inferential framework, but its expression in behavior depends on task demands and the temporal dynamics of sensory processing.

Taken together, these findings support a unified account in which multisensory perception is governed by a stable causal inference structure that is preserved across spatial contexts and levels of visual reliability. Variability in behavior arises not from changes in the binding mechanism itself, but from differences in sensory uncertainty and the brain's ability to estimate it. This framework provides a principled foundation for understanding perception under extreme uncertainty and offers a computational basis for designing systems that support adaptive sensory integration in real-world environments.

Chapter 6

GENERAL CONCLUSION

The central question of this dissertation is: how does the brain make sense of the world when sensory information becomes unreliable? To address this, I examined multisensory inference across three forms of visual uncertainty that differ in structure and learnability: First, chronic pathological uncertainty in individuals with low vision; Second, the physiological blind spot, where visual input is absent due to the lack of photoreceptors; and finally, experimentally imposed uncertainty in immersive virtual reality, where visual reliability is manipulated in real time.

Across these contexts, a consistent pattern emerges: multisensory perception is governed by a stable causal inference structure, but its effectiveness depends critically on how well the brain can estimate sensory reliability.

First, in low vision (**Chapter 3**), visual input is degraded in a spatially heterogeneous and continuously varying manner. Under these conditions, illusion strength is not enhanced, and model performance becomes weaker and more variable, particularly at low-visibility locations. This contrasts with sensory compensation accounts proposing that crossmodal interactions strengthen following vision loss, potentially leading to enhanced multisensory perception and therefore stronger illusion susceptibility. Instead, we found that perceiving the illusion depends critically on reliable visual input, as illusion strength increases with local visual reliability. Importantly, this does not indicate that participants adopt fundamentally non-Bayesian inference strategies. Rather, behavior becomes increasingly noisy when visual detection is unreliable, making it difficult for standard BCI models to fully capture the observed response patterns. In addition, this study was optimized for broad spatial sampling across many retinal locations rather than exhaustive condition sampling for computational model fitting, so fitted parameters should be interpreted cautiously. However, when the analysis is restricted to high-visibility locations, the model produced strong fits for both the low vision and sighted control groups. This suggests that multisensory cue integration and segregation remain broadly consistent with Bayesian inference when visual signals are sufficiently reliable. Together, these findings indicate that chronic visual impairment does not fundamentally alter causal inference, but instead limits the stability and reliability of the sensory evidence on

which inference operates.

Second, in the blind spot (**Chapter 4**), visual input is absent but in a highly structured and lifelong predictable manner. Here, observers exhibit near-optimal Bayesian Causal Inference: illusory percepts remain precise, and BCI model fits are strong. This suggests that the brain has learned, over the lifespan, the statistical properties of this uncertainty due to the blind spot and compensates by adjusting sensory likelihoods and priors. In contrast to low vision, where uncertainty is unstable and difficult to estimate, the blind spot represents a case where uncertainty is fully internalized, allowing the inference process to operate efficiently.

Finally, in virtual reality (**Chapter 5**), visual uncertainty is externally imposed and dynamically manipulated. Observers continue to follow the same computational principles, but their behavior becomes less optimal as visual reliability decreases. Although BCI still captures behavior well relative to alternative models, its predictive power declines under low reliability. This reflects a limitation in rapidly estimating and recalibrating unfamiliar, time-varying uncertainty, even when the underlying inference structure remains unchanged.

Taken together, these findings reveal both consistency and divergence across conditions. The consistency lies in the preservation of the causal inference framework: the brain does not globally increase binding tendency or switch to a different computational strategy under degraded vision. The divergence lies in how effectively this framework can be applied. When uncertainty is stable and well learned, as in the blind spot, inference is Bayes-optimal. When uncertainty is unstable, heterogeneous, or intermittently absent, as in low vision and experimental manipulation in virtual reality, inference becomes less efficient.

This pattern reveals an important boundary condition of the Bayesian Causal Inference (BCI) framework. BCI is fundamentally a decision model that determines whether multiple sensory signals should be integrated or treated as separate causes. Critically, this computation requires the presence of at least two concurrently available sensory estimates. When one modality becomes effectively absent, the problem is no longer one of multisensory causal inference, but instead reduces to a predominantly unimodal decision process.

Our findings suggest that this distinction is essential for interpreting behavior under degraded sensory conditions. In the physiological blind spot, visual uncertainty is stable, spatially consistent, and well learned across the lifespan. Although visual

input is absent at the blind spot itself, the surrounding visual context remains reliable, allowing multisensory inference to operate in a predictable and near-optimal manner. In contrast, low vision introduces spatially heterogeneous and unstable sensory degradation. At severely impaired retinal locations, participants are effectively locally “blind”, such that perception is dominated by auditory information with little or no usable visual estimate available for integration. Under these conditions, there may no longer be two meaningful sensory signals for the brain to evaluate within the BCI framework.

A similar consideration motivated the design and analysis of the immersive virtual reality study. We restricted computational modeling to the front-facing region, where visual information remained concurrently available with tactile cues and participants could plausibly perform multisensory integration. Side and rear regions were excluded from model fitting because participant behavior in those areas was more consistent with a sequential strategy, where tactile feedback first guided orienting behavior before vision was used to refine the final response. In these conditions, the task no longer reflected simultaneous multisensory estimation assumed by standard BCI models.

Importantly, however, even within the front-facing region where multisensory integration remained possible, model fits became systematically poorer under low visual reliability. This suggests that degraded performance cannot be explained solely by the absence of concurrent sensory estimates. Rather, when visual signals become highly unreliable, noisy, or difficult to characterize, participants may struggle to form stable estimates of sensory reliability itself. As a result, behavior deviates from model predictions not necessarily because causal inference breaks down, but because the underlying sensory evidence no longer conforms to the assumptions of stable, continuously distributed uncertainty embedded in standard BCI frameworks.

Taken together, these findings suggest that the primary challenge is not simply increased sensory noise, but whether reliable concurrent sensory estimates exist in the first place. When signals are stable and jointly available, multisensory inference remains remarkably robust. But when one modality becomes effectively unavailable, highly unstable, or only sequentially informative, behavior shifts away from the conditions under which causal inference models are designed to operate.

More broadly, this dissertation supports a more constrained interpretation of multisensory integration under uncertainty. The brain does not indiscriminately increase sensory binding when vision is degraded. Instead, multisensory inference appears

to depend critically on the availability, stability, and temporal coexistence of sensory evidence. Perception is therefore not only about integrating noisy signals, but also about determining whether there are multiple reliable signals available to integrate at all.

This perspective also points to important directions for future work. Rather than focusing solely on the magnitude of sensory noise, future research should characterize the structure and dynamics of reliability itself: how it varies across space, how it changes over time, and how the brain represents conditions where signals are intermittently absent rather than continuously noisy. Understanding how reliability is learned, updated, and sometimes fails to stabilize may be key to explaining when multisensory inference remains Bayes-optimal and when it breaks down.

In this sense, perception can be understood as a process of continuous calibration. Supporting this calibration process, particularly in conditions where sensory reliability is unstable or degraded, may provide a principled direction for developing interventions in populations with sensory impairment.

In conclusion, this dissertation shows that the brain does not undergo a global reorganization in response to vision loss. Instead, it preserves a common inference structure while adapting to local constraints on sensory reliability, demonstrating that perception is fundamentally an exercise not only in integrating information, but in managing uncertainty about that information.

Bibliography

- [1] Luigi Acerbi, Kalpana Dokka, Dora E. Angelaki, and Wei Ji Ma. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLOS Computational Biology*, 14 (7):e1006110, July 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006110. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006110>. Publisher: Public Library of Science.
- [2] J. L. Ackrill and J. L. Ackrill, editors. *A New Aristotle Reader*. Oxford University Press, Oxford, New York, October 1987. ISBN 978-0-19-875061-1.
- [3] David Alais and David Burr. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262, February 2004. ISSN 0960-9822. doi: 10.1016/j.cub.2004.01.029. URL [https://www.cell.com/current-biology/abstract/S0960-9822\(04\)00043-0](https://www.cell.com/current-biology/abstract/S0960-9822(04)00043-0).
- [4] Máté Aller and Uta Noppeney. To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLOS Biology*, 17 (4):e3000210, April 2019. ISSN 1545-7885. doi: 10.1371/journal.pbio.3000210. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000210>.
- [5] Oscar J. Ariza N., Markus Lange, Frank Steinicke, and Gerd Bruder. Vibrotactile assistance for user guidance towards selection targets in VR and the cognitive resources involved. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 95–98, March 2017. doi: 10.1109/3DUI.2017.7893323. URL <https://ieeexplore.ieee.org/document/7893323>.
- [6] Stacey Aston, Marko Nardini, and Ulrik Beierholm. Different types of uncertainty in multisensory perceptual decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1886):20220349, 2023. ISSN 0962-8436. doi: 10.1098/rstb.2022.0349. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC10404920/>.
- [7] Jeroen Atsma, Femke Maij, Mathieu Koppen, David E. Irwin, and W. Pieter Medendorp. Causal Inference for Spatial Constancy across Saccades. *PLOS Computational Biology*, 12(3):e1004766, March 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004766. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004766>. Publisher: Public Library of Science.
- [8] Paul Bach-y Rita. Tactile sensory substitution studies. *Annals of the New York Academy of Sciences*, 1013:83–91, May 2004. ISSN 0077-8923. doi: 10.1196/annals.1305.006.

- [9] Anton L. Beer, Tina Plank, and Mark W. Greenlee. Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental Brain Research*, 213(2):299–308, September 2011. ISSN 1432-1106. doi: 10.1007/s00221-011-2715-y. URL <https://doi.org/10.1007/s00221-011-2715-y>.
- [10] Ulrik R. Beierholm, Steven R. Quartz, and Ladan Shams. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, 9(5):23, May 2009. ISSN 1534-7362. doi: 10.1167/9.5.23. URL <https://doi.org/10.1167/9.5.23>.
- [11] Almut Braun. The sensory compensation hypothesis. In Almut Braun, editor, *The Speaker Identification Ability of Blind and Sighted Listeners: An Empirical Investigation*, pages 27–29. Springer Fachmedien, Wiesbaden, 2016. ISBN 978-3-658-15198-0. doi: 10.1007/978-3-658-15198-0_1. URL https://doi.org/10.1007/978-3-658-15198-0_1.
- [12] Simona Caraiman, Otilia Zvoristeanu, Adrian Burlacu, and Paul Herghelegiu. Stereo Vision Based Sensory Substitution for the Visually Impaired. *Sensors*, 19(12):2771, January 2019. doi: 10.3390/s19122771. URL <https://www.mdpi.com/1424-8220/19/12/2771>. Number: 12.
- [13] Ailene Chan. chanyca/blindspot-multisensory: Full Release with New Analyses, November 2025. URL <https://doi.org/10.5281/zenodo.15208160>.
- [14] Ailene Chan. chanyca/illusory-mapping: Initial release, December 2025. URL <https://zenodo.org/records/18049129>.
- [15] Ailene Y. C. Chan, Noelle R. B. Stiles, Carmel A. Levitan, Daw-An Wu, Armand R. Tanguay, and Shinsuke Shimojo. Bayesian Causal Inference Accounts for Multisensory Filling-In at the Blind Spot. *bioRxiv*, page 2024.11.15.623713, January 2025. doi: 10.1101/2024.11.15.623713. URL <http://biorxiv.org/content/early/2025/11/20/2024.11.15.623713.abstract>.
- [16] Chun Chang, Erlei Wang, Jiajia Yang, Xiaotong Luan, Aijun Wang, and Ming Zhang. Differences in eccentricity for sound-induced flash illusion in four visual fields. *Perception*, 52(1):56–73, January 2023. ISSN 0301-0066. doi: 10.1177/03010066221136670. URL <https://doi.org/10.1177/03010066221136670>.
- [17] Daniel-Robert Chebat, Constant Rainville, Ron Kupers, and Maurice Ptito. Tactile-’visual’ acuity of the tongue in early blind individuals. *Neuroreport*, 18(18):1901–1904, December 2007. ISSN 0959-4965. doi: 10.1097/WNR.0b013e3282f2a63.

- [18] Yi-Chuan Chen, Daphne Maurer, Terri L. Lewis, Charles Spence, and David I. Shore. Central–peripheral differences in audiovisual and visuotactile event perception. *Attention, Perception, & Psychophysics*, 79(8):2552–2563, November 2017. ISSN 1943-393X. doi: 10.3758/s13414-017-1396-4. URL <https://doi.org/10.3758/s13414-017-1396-4>.
- [19] Seungmoon Choi and Katherine J. Kuchenbecker. Vibrotactile Display: Perception, Technology, and Applications. *Proceedings of the IEEE*, 101(9):2093–2104, September 2013. ISSN 1558-2256. doi: 10.1109/JPROC.2012.2221071. URL <https://ieeexplore.ieee.org/document/6353870>.
- [20] Roger W. Cholewiak, J. Christopher Brill, and Anja Schwab. Vibrotactile localization on the abdomen: effects of place and space. *Perception & Psychophysics*, 66(6):970–987, August 2004. ISSN 0031-5117. doi: 10.3758/bf03194989.
- [21] L. G. Cohen, P. Celnik, A. Pascual-Leone, B. Corwell, L. Falz, J. Dambrosia, M. Honda, N. Sadato, C. Gerloff, M. D. Catalá, and M. Hallett. Functional relevance of cross-modal plasticity in blind humans. *Nature*, 389(6647):180–183, September 1997. ISSN 0028-0836. doi: 10.1038/38278.
- [22] Olivier Collignon, Patrice Voss, Maryse Lassonde, and Franco Lepore. Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, 192(3):343–358, January 2009. ISSN 1432-1106. doi: 10.1007/s00221-008-1553-z. URL <https://doi.org/10.1007/s00221-008-1553-z>.
- [23] Samantha I. Cunningham, James D. Weiland, Pinglei Bao, Gilberto Raul Lopez-Jaime, and Bosco S. Tjan. Correlation of vision loss with tactile-evoked V1 responses in retinitis pigmentosa. *Vision Research*, 111(Pt B):197–207, June 2015. ISSN 1878-5646. doi: 10.1016/j.visres.2014.10.015.
- [24] Victor Adriel de Jesus Oliveira, Luca Brayda, Luciana Nedel, and Anderson Maciel. Designing a Vibrotactile Head-Mounted Display for Spatial Awareness in 3D Spaces. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1409–1417, April 2017. ISSN 1941-0506. doi: 10.1109/TVCG.2017.2657238. URL <https://ieeexplore.ieee.org/document/7829406/authors>.
- [25] Frédéric Devinck and Kenneth Knoblauch. Central mechanisms of perceptual filling-in. *Current Opinion in Behavioral Sciences*, 30:135–140, December 2019. ISSN 2352-1546. doi: 10.1016/j.cobeha.2019.08.003. URL <https://www.sciencedirect.com/science/article/pii/S2352154619300427>.
- [26] Kalpana Dokka, Hyeshin Park, Michael Jansen, Gregory C. DeAngelis, and Dora E. Angelaki. Causal inference accounts for heading perception in the

- presence of object motion. *Proceedings of the National Academy of Sciences*, 116(18):9060–9065, April 2019. doi: 10.1073/pnas.1820373116. URL <https://www.pnas.org/doi/full/10.1073/pnas.1820373116>. Publisher: Proceedings of the National Academy of Sciences.
- [27] Andreas Aldogan Eklund and Miralem Helmeffalk. Seeing through touch: a conceptual framework of visual-tactile interplay. *Journal of Product & Brand Management*, 27(5):498–513, January 2018. ISSN 1061-0421. doi: 10.1108/JPBM-07-2017-1520. URL <https://doi.org/10.1108/JPBM-07-2017-1520>.
- [28] Linda R. Elliott and Elizabeth S. Redden. Reducing Workload: A Multisensory Approach. In *Designing Soldier Systems*. CRC Press, 2012. Num Pages: 28.
- [29] Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, January 2002. ISSN 1476-4687. doi: 10.1038/415429a. URL <https://doi.org/10.1038/415429a>.
- [30] Arnaud Falchier, Simon Clavagnier, Pascal Barone, and Henry Kennedy. Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. *Journal of Neuroscience*, 22(13):5749–5759, July 2002. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.22-13-05749.2002. URL <https://www.jneurosci.org/content/22/13/5749>.
- [31] Niall Gavin, Rebecca J. Hirst, and David P. McGovern. The magnitude of the sound-induced flash illusion does not increase monotonically as a function of visual stimulus eccentricity. *Attention, Perception, & Psychophysics*, 84(5): 1689–1698, July 2022. ISSN 1943-393X. doi: 10.3758/s13414-022-02493-4. URL <https://doi.org/10.3758/s13414-022-02493-4>.
- [32] Monica Gori, Giulio Sandini, Cristina Martinoli, and David C. Burr. Impairment of auditory spatial localization in congenitally blind human subjects. *Brain: A Journal of Neurology*, 137(Pt 1):288–293, January 2014. ISSN 1460-2156. doi: 10.1093/brain/awt311.
- [33] David C Knill and Whitman Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.
- [34] Kenneth J. Kokjer. The information capacity of the human fingertip. *IEEE Transactions on Systems, Man, & Cybernetics*, 17(1):100–102, 1987. ISSN 0018-9472. doi: 10.1109/TSMC.1987.289337.
- [35] Hidehiko Komatsu. The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience*, 7(3):220–231, March 2006. ISSN 1471-0048. doi: 10.1038/nrn1869. URL <https://www.nature.com/articles/nrn1869>.

- [36] Daniel P. Kumpik, Helen E. Roberts, Andrew J. King, and Jennifer K. Bizley. Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision*, 14(7):12, June 2014. ISSN 1534-7362. doi: 10.1167/14.7.12. URL <https://doi.org/10.1167/14.7.12>.
- [37] Ron Kupers and Maurice Ptito. Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neuroscience & Biobehavioral Reviews*, 41:36–52, April 2014. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2013.08.001. URL <https://www.sciencedirect.com/science/article/pii/S0149763413001917>.
- [38] Konrad P. Körding, Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B. Tenenbaum, and Ladan Shams. Causal Inference in Multisensory Perception. *PLOS ONE*, 2(9):e943, September 2007. ISSN 1932-6203. doi: 10.1371/journal.pone.0000943. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000943>.
- [39] Tayfun Lloyd-Esenkaya, Vanessa Lloyd-Esenkaya, Eamonn O’Neill, and Michael J. Proulx. Multisensory inclusive design with sensory substitution. *Cognitive Research: Principles and Implications*, 5(1):37, August 2020. ISSN 2365-7464. doi: 10.1186/s41235-020-00240-7. URL <https://doi.org/10.1186/s41235-020-00240-7>.
- [40] Granit Luzhnica, Sebastian Stein, Eduardo Veas, Viktoria Pammer, John Williamson, and Roderick Murray Smith. Personalising vibrotactile displays through perceptual sensitivity adjustment. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC ’17, pages 66–73, New York, NY, USA, September 2017. Association for Computing Machinery. ISBN 978-1-4503-5188-1. doi: 10.1145/3123021.3123029. URL <https://doi.org/10.1145/3123021.3123029>.
- [41] Neil A. Macmillan and Howard L. Kaplan. Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1):185–199, 1985. ISSN 1939-1455. doi: 10.1037/0033-2909.98.1.185. Place: US.
- [42] Matteo Marucci, Gianluca Di Flumeri, Gianluca Borghini, Nicolina Sciaraffa, Michele Scandola, Enea Francesco Pavone, Fabio Babiloni, Viviana Betti, and Pietro Aricò. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports*, 11(1):4831, March 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-84196-8. URL <https://www.nature.com/articles/s41598-021-84196-8>.
- [43] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*, page 105, 1974.

- [44] Harry McGurk and John Macdonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, December 1976. ISSN 1476-4687. doi: 10.1038/264746a0. URL <https://www.nature.com/articles/264746a0>. Number: 5588.
- [45] P.B.L. Meijer. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121, February 1992. ISSN 1558-2531. doi: 10.1109/10.121642. Conference Name: IEEE Transactions on Biomedical Engineering.
- [46] Liam J. Norman and Lore Thaler. Retinotopic-like maps of spatial sound in primary ‘visual’ cortex of blind human echolocators. *Proceedings of the Royal Society B: Biological Sciences*, 286(1912):20191910, October 2019. doi: 10.1098/rspb.2019.1910. URL <https://royalsocietypublishing.org/doi/full/10.1098/rspb.2019.1910>.
- [47] Valeria Occelli, Charles Spence, and Massimiliano Zampini. Auditory, tactile, and audiotactile information processing following visual deprivation. *Psychological Bulletin*, 139(1):189–212, January 2013. ISSN 1939-1455. doi: 10.1037/a0028416.
- [48] Brian Odegaard and Ladan Shams. The Brain’s Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*, 27(4):583–591, April 2016. ISSN 0956-7976. doi: 10.1177/0956797616628860. URL <https://doi.org/10.1177/0956797616628860>.
- [49] Brian Odegaard, David R. Wozny, and Ladan Shams. The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, 614:24–28, February 2016. ISSN 0304-3940. doi: 10.1016/j.neulet.2015.12.039. URL <https://www.sciencedirect.com/science/article/pii/S030439401530327X>.
- [50] Brian Odegaard, David R. Wozny, and Ladan Shams. A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ*, 5:e3143, April 2017. ISSN 2167-8359. doi: 10.7717/peerj.3143. URL <https://peerj.com/articles/3143>.
- [51] Emily Orchard-Mills, Erik Van der Burg, and David Alais. Crossmodal Correspondence Between Auditory Pitch and Visual Elevation Affects Temporal Ventriloquism. *Perception*, 45(4):409–424, April 2016. ISSN 0301-0066. doi: 10.1177/0301006615622320. URL <https://doi.org/10.1177/0301006615622320>.
- [52] J Paillard. Fast and slow feedback loops for the visual correction of spatial errors in a pointing task: a reappraisal. *Canadian Journal of Physiology and Pharmacology*, 74(4):401–417, April 1996. ISSN 0008-4212, 1205-7541. doi: 10.1139/y96-033. URL <http://www.nrcresearchpress.com/doi/10.1139/y96-033>.

- [53] Mark Paterson. Hearing Gloves and Seeing Tongues? Disability, Sensory Substitution and the Origins of the Neuroplastic Subject. *Body & Society*, page 1357034X211008235, July 2021. ISSN 1357-034X. doi: 10.1177/1357034X211008235. URL <https://doi.org/10.1177/1357034X211008235>.
- [54] Florian Perdreau, James R. H. Cooke, Mathieu Koppen, and W. Pieter Medendorp. Causal inference for spatial constancy across whole body motion. *Journal of Neurophysiology*, 121(1):269–284, January 2019. ISSN 0022-3077. doi: 10.1152/jn.00473.2018. URL <https://journals.physiology.org/doi/full/10.1152/jn.00473.2018>. Publisher: American Physiological Society.
- [55] Megan A. K. Peters, Ling-Qi Zhang, and Ladan Shams. The material-weight illusion is a Bayes-optimal percept under competing density priors. *PeerJ*, 6:e5760, October 2018. ISSN 2167-8359. doi: 10.7717/peerj.5760. URL <https://peerj.com/articles/5760>. Publisher: PeerJ Inc.
- [56] Maurice Ptito, Isabelle Matteau, Arthur Zhi Wang, Olaf B. Paulson, Hartwig R. Siebner, and Ron Kupers. Crossmodal recruitment of the ventral visual stream in congenital blindness. *Neural Plasticity*, 2012:304045, 2012. ISSN 1687-5443. doi: 10.1155/2012/304045.
- [57] Juan F. Quinones, Tommaso Pavan, Xinyang Liu, Christiane M. Thiel, Axel Heep, and Andrea Hildebrandt. Fiber tracing and microstructural characterization among audiovisual integration brain regions in neonates compared with young adults. *NeuroImage*, 254:119141, July 2022. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2022.119141. URL <https://www.sciencedirect.com/science/article/pii/S1053811922002580>.
- [58] Lara Rahal, Jongeun Cha, and Abdulmotaleb El Saddik. Continuous tactile perception for vibrotactile displays. In *2009 IEEE International Workshop on Robotic and Sensors Environments*, pages 86–91, November 2009. doi: 10.1109/ROSE.2009.5355986.
- [59] Dobromir Rahnev and Rachel N. Denison. Suboptimality in perceptual decision making. *The Behavioral and Brain Sciences*, 41:e223, February 2018. ISSN 1469-1825. doi: 10.1017/S0140525X18000936.
- [60] Vilayanur S. Ramachandran. Blind Spots. *Scientific American*, 266(5):86–91, May 1992. doi: <https://doi.org/10.1038/scientificamerican0592-86>. URL <https://www.scientificamerican.com/article/blind-spots/>.
- [61] Rajani Raman and Sandip Sarkar. Predictive Coding: A Possible Explanation of Filling-In at the Blind Spot. *PLOS ONE*, 11(3):e0151194, March 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0151194. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0151194>.

- [62] Michael D. Richards, Herbert C. Goltz, and Agnes M. F. Wong. Temporal Ventriloquism Reveals Intact Audiovisual Temporal Integration in Amblyopia. *Investigative Ophthalmology & Visual Science*, 59(2):637–643, February 2018. ISSN 1552-5783. doi: 10.1167/iovs.17-22613. URL <http://iovs.arvojournals.org/article.aspx?articleid=2671732>.
- [63] Tim Rohe and Uta Noppeney. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biology*, 13(2): e1002073, February 2015. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002073. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002073>.
- [64] N. Sadato, A. Pascual-Leone, J. Grafman, V. Ibañez, M. P. Deiber, G. Dold, and M. Hallett. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, 380(6574):526–528, April 1996. ISSN 0028-0836. doi: 10.1038/380526a0.
- [65] Majed Samad, Albert Jin Chung, and Ladan Shams. Perception of Body Ownership Is Driven by Bayesian Sensory Inference. *PLOS ONE*, 10(2): e0117178, February 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0117178. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117178>.
- [66] Meike Scheller, Michael J. Proulx, Michelle de Haan, Annegret Dahlmann-Noor, and Karin Petrini. Late- but not early-onset blindness impairs the development of audio-haptic multisensory integration. *Developmental Science*, 24(1):e13001, 2021. ISSN 1467-7687. doi: 10.1111/desc.13001. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.13001>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.13001>.
- [67] Irene Senna, Elena Andres, Ayelet McKyton, Itay Ben-Zion, Ehud Zohary, and Marc O. Ernst. Development of multisensory integration following prolonged early-onset visual deprivation. *Current Biology*, 31(21): 4879–4885.e6, November 2021. ISSN 0960-9822. doi: 10.1016/j.cub.2021.08.060. URL [https://www.cell.com/current-biology/abstract/S0960-9822\(21\)01196-9](https://www.cell.com/current-biology/abstract/S0960-9822(21)01196-9).
- [68] Ladan Shams. Early Integration and Bayesian Causal Inference in Multisensory Perception, 2012. URL <http://europepmc.org/books/NBK92847>. ISBN: 9781439812174 Series: Frontiers in Neuroscience.
- [69] Ladan Shams and Ulrik Beierholm. Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, 137: 104619, June 2022. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2022.104619. URL <https://www.sciencedirect.com/science/article/pii/S0149763422001087>.

- [70] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. What you see is what you hear. *Nature*, 408(6814):788–788, December 2000. ISSN 1476-4687. doi: 10.1038/35048669. URL <https://www.nature.com/articles/35048669>. Number: 6814.
- [71] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Visual illusion induced by sound. *Cognitive Brain Research*, 14(1):147–152, June 2002. ISSN 0926-6410. doi: 10.1016/S0926-6410(02)00069-1. URL <https://www.sciencedirect.com/science/article/pii/S0926641002000691>.
- [72] Ladan Shams, Wei Ji Ma, and Ulrik Beierholm. Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17):1923–1927, November 2005. ISSN 0959-4965. doi: 10.1097/01.wnr.0000187634.68504.bb.
- [73] Charles Spence. The Skin as a Medium for Sensory Substitution. *Multi-sensory Research*, 27(5-6):293–312, January 2014. ISSN 2213-4808, 2213-4794. doi: 10.1163/22134808-00002452. URL https://brill.com/view/journals/msr/27/5-6/article-p293_3.xml.
- [74] Noelle R. B. Stiles, Monica Li, Carmel A. Levitan, Yukiyasu Kamitani, and Shinsuke Shimojo. What you saw is what you will hear: Two new illusions with audiovisual postdictive effects. *PLOS ONE*, 13(10): e0204217, October 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0204217. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204217>.
- [75] Michele Tagliabue and Joseph McIntyre. A modular theory of multisensory integration for motor control. *Frontiers in Computational Neuroscience*, 8, January 2014. ISSN 1662-5188. doi: 10.3389/fncom.2014.00001. URL <https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2014.00001/full>.
- [76] Stefano Targher, Valeria Occelli, and Massimiliano Zampini. Audiovisual integration in low vision individuals. *Neuropsychologia*, 50(5):576–582, April 2012. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2011.10.021. URL <https://www.sciencedirect.com/science/article/pii/S0028393211004969>.
- [77] Peter I. Terrence, J. Christopher Brill, and Richard D. Gilson. Body Orientation and the Perception of Spatial Auditory and Tactile Cues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(17):1663–1667, September 2005. ISSN 1071-1813. doi: 10.1177/154193120504901735. URL <https://doi.org/10.1177/154193120504901735>.
- [78] Jan Van Erp. Presenting directions with a vibrotactile torso display. *Ergonomics*, 48(3):302–313, February 2005. ISSN 0014-0139. doi: 10.1080/0014013042000327670.

- [79] Jan B. F. van Erp. Absolute localization of vibrotactile stimuli on the torso. *Perception & Psychophysics*, 70(6):1016–1023, August 2008. ISSN 0031-5117. doi: 10.3758/pp.70.6.1016.
- [80] Tiziana Vercillo, David Burr, and Monica Gori. Early visual deprivation severely compromises the auditory sense of space in congenitally blind children. *Developmental Psychology*, 52(6):847–853, June 2016. ISSN 1939-0599. doi: 10.1037/dev0000103.
- [81] Erik Verhaar, Wijbrand Pieter Medendorp, Sabine Hunnius, and Janny C. Stapel. Bayesian causal inference in visuotactile integration in children and adults. *Developmental Science*, 25(3):e13184, 2022. ISSN 1467-7687. doi: 10.1111/desc.13184. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.13184>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.13184>.
- [82] Catherine Y. Wan, Amanda G. Wood, David C. Reutens, and Sarah J. Wilson. Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia*, 48(1):344–348, January 2010. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2009.08.016. URL <https://www.sciencedirect.com/science/article/pii/S0028393209003340>.
- [83] Jamie Ward and Peter Meijer. Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, 19(1):492–500, March 2010. ISSN 1053-8100. doi: 10.1016/j.concog.2009.10.006. URL <https://www.sciencedirect.com/science/article/pii/S1053810009001718>.
- [84] Rimona S. Weil and Geraint Rees. A new taxonomy for perceptual filling-in. *Brain Research Reviews*, 67(1):40–55, June 2011. ISSN 0165-0173. doi: 10.1016/j.brainresrev.2010.10.004. URL <https://www.sciencedirect.com/science/article/pii/S0165017310001177>.
- [85] Michael Wong, Vishi Gnanakumaran, and Daniel Goldreich. Tactile Spatial Acuity Enhancement in Blindness: Evidence for Experience-Dependent Mechanisms. *Journal of Neuroscience*, 31(19):7028–7037, May 2011. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.6461-10.2011. URL <https://www.jneurosci.org/content/31/19/7028>.
- [86] R. S. Woodworth. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, 3(3):i–114, 1899. ISSN 0096-9753. doi: 10.1037/h0092992.
- [87] David R. Wozny, Ulrik R. Beierholm, and Ladan Shams. Human trimodal perception follows optimal statistical inference. *Journal of Vision*, 8(3):24, March 2008. ISSN 1534-7362. doi: 10.1167/8.3.24. URL <https://doi.org/10.1167/8.3.24>.

- [88] David R. Wozny, Ulrik R. Beierholm, and Ladan Shams. Probabil-
ity Matching as a Computational Strategy Used in Perception. *PLOS*
Computational Biology, 6(8):e1000871, August 2010. ISSN 1553-7358.
doi: 10.1371/journal.pcbi.1000871. URL [https://journals.plos.org/
ploscompbiol/article?id=10.1371/journal.pcbi.1000871](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000871).
- [89] Haocheng Zhu, Ulrik Beierholm, and Ladan Shams. BCI Toolbox: An
open-source python package for the Bayesian causal inference model. *PLOS*
Computational Biology, 20(7):e1011791, July 2024. ISSN 1553-7358.
doi: 10.1371/journal.pcbi.1011791. URL [https://journals.plos.org/
ploscompbiol/article?id=10.1371/journal.pcbi.1011791](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011791).
- [90] Haocheng Zhu, Ulrik Beierholm, and Ladan Shams. The overlooked role
of unisensory precision in multisensory research. *Current Biology*, 34(6):
R229–R231, March 2024. ISSN 0960-9822. doi: 10.1016/j.cub.2024.
01.057. URL [https://www.cell.com/current-biology/abstract/
S0960-9822\(24\)00094-0](https://www.cell.com/current-biology/abstract/S0960-9822(24)00094-0).

Appendix A

SUPPLEMENTAL INFORMATION FOR CHAPTER III

Beep Detection Task and Auditory Parameter Validation

This task was performed by a subset of 16 newly recruited age-matched controls following the completion of data collection in the low-vision cohort. The primary objective was to empirically validate the auditory likelihood parameter (σ_A) used in our BCI modeling.

Methods

The visual and auditory stimuli in this task were identical to that in the main experiment (Visual Flash Detection Task, Illusory Double Flash Task). Visual stimuli were (circular white discs (2° in diameter) presented on a black background at 100% contrast for one frame (16.7 ms at 60 Hz). Auditory stimuli were pure tones (800 Hz, 7 ms duration). Six audiovisual combinations of flashes (0 or 1) and beeps (0, 1, or 2) were tested. Stimulus timing followed the same parameters as described in the main text (see Fig. 1B–E).

Parameter Estimation

To derive σ_A , we fit a unisensory Gaussian observer model to the numerosity responses. To ensure the estimation was not biased by visual parameters, we utilized a "flat-visual" BCI framework where:

- $\sigma_V = 4000$: Fixed to ensure the visual likelihood provided no numerosity information.
- $p_{common} = 0$: Fixed to assume independent causes, effectively bypassing the fusion logic of the BCI model.
- $\mu_P = 0, \sigma_P = 4000$: Fixed to ensure a flat, uninformative prior
- σ_A : Remained the only free parameter.

Results

The empirical estimate for auditory reliability in our control subset ($\sigma_A = 0.19 \pm 0.08$) confirms that auditory sensitivity in our age-matched participants is highly

consistent with the $\sigma_A = 0.2$ value established in previous studies using identical stimuli [15]. This high precision is further supported by the near-ceiling performance observed in our beep-detection task (Fig. A.1). Given this cross-study consistency and the minimal variance observed in the control subset, we fixed σ_A at 0.2 across both experimental groups. This constraint reduces the model’s complexity and facilitates a direct, parsimonious comparison of visual reliability (σ_V) and causal priors (P_{common}) between the Sighted Control and Low Vision populations.

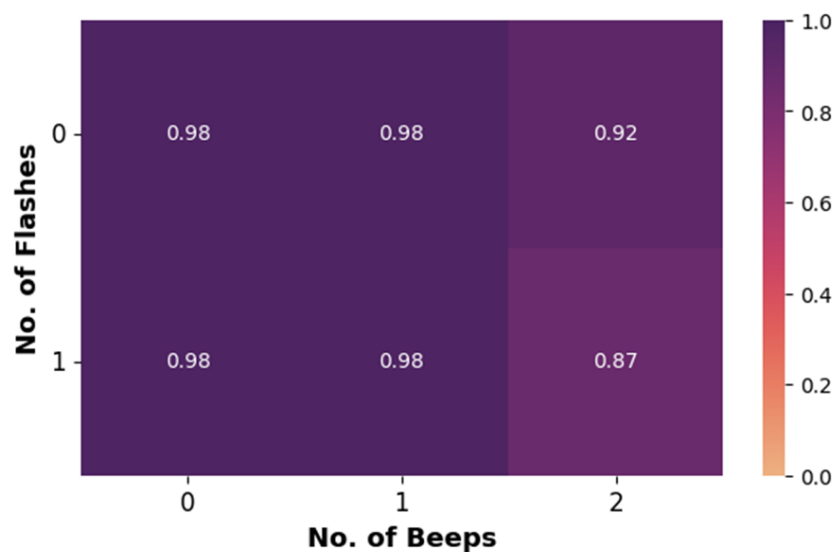


Figure A.1: **Mean Accuracy of Perceiving Beeps (N=16).** Heatmaps show accuracy in reporting the true number of beeps (0/1/2) across conditions. Performance is near ceiling (1.0) across conditions, suggesting that number of flashes present has little influence on beep detection accuracy.

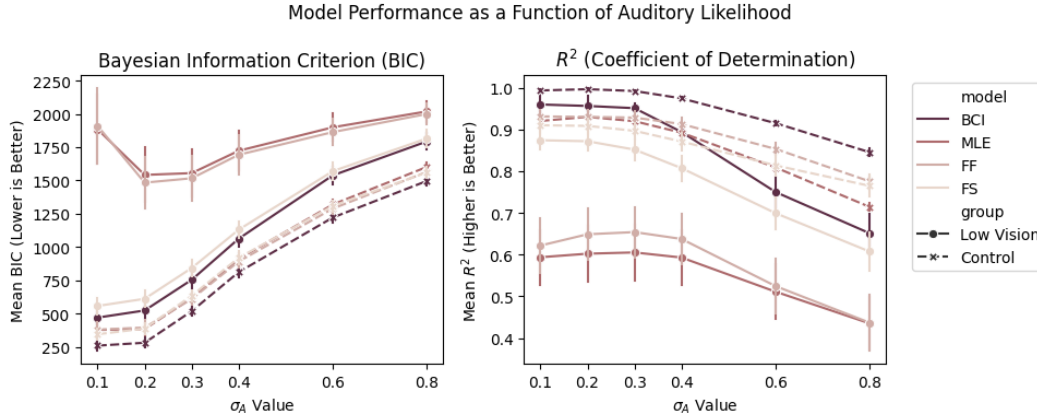


Figure A.2: **Sensitivity Analysis for different σ_A values.** To validate the selection of the fixed auditory likelihood parameter (σ_A), we performed a parameter sweep across a range of values (0.1, 0.2, 0.3, 0.4, 0.6, 0.8). By visual inspection, it was apparent that 0.1-0.3 yielded the best model performance (low BIC scores and high R^2). We performed Wilcoxon signed-rank tests to compare BIC and R^2 between 0.1 and 0.2, and 0.2 and 0.3. After correcting for multiple comparisons (Bonferroni), Model $_{\sigma_A=0.3}$ performs significantly worse than Model $_{\sigma_A=0.2}$. Among the control group, while Model $_{\sigma_A=0.1}$ has significantly lower BIC scores than Model $_{\sigma_A=0.2}$, there is no meaningful difference in R^2 .

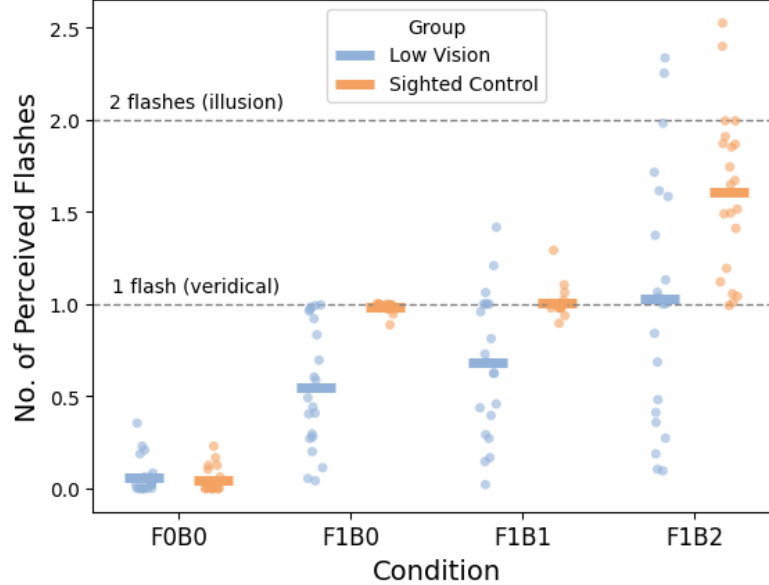


Figure A.3: **Auditory influence on flash perception.** Both groups show auditory influence on the number of flashes perceived. Each dot represents one participant, horizontal bar represents the group average.

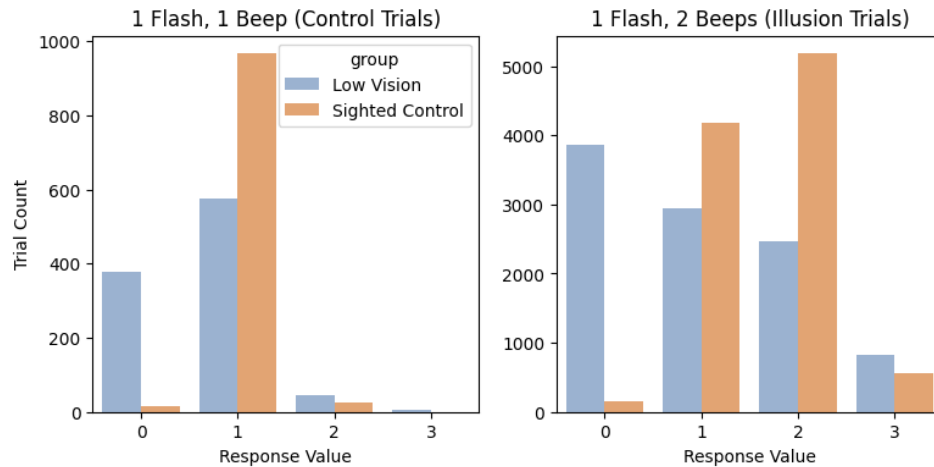


Figure A.4: **Trial-level response distributions for Sighted Control and Low Vision groups.** Bars represent the total count of trials for each response value (number of flashes perceived). Left: 1 Flash, 1 Beep (Control Trials) – While Sighted Controls show high detection consistency (peaking at 1), the Low Vision group exhibits a substantial number of “misses” (0 flashes), indicating a lower sensory detection threshold. Right: 1 Flash, 2 Beeps (Illusion Trials) – Inaccuracy in Sighted Controls is driven by a shift toward reporting 2 flashes (the sound-induced flash illusion). In the Low Vision group, inaccuracy is split between “misses” (0) and “illusions” (2). This bimodal error distribution explains why the Low Vision group’s mean perceived flashes can remain near 1.0 despite significantly lower accuracy than controls.

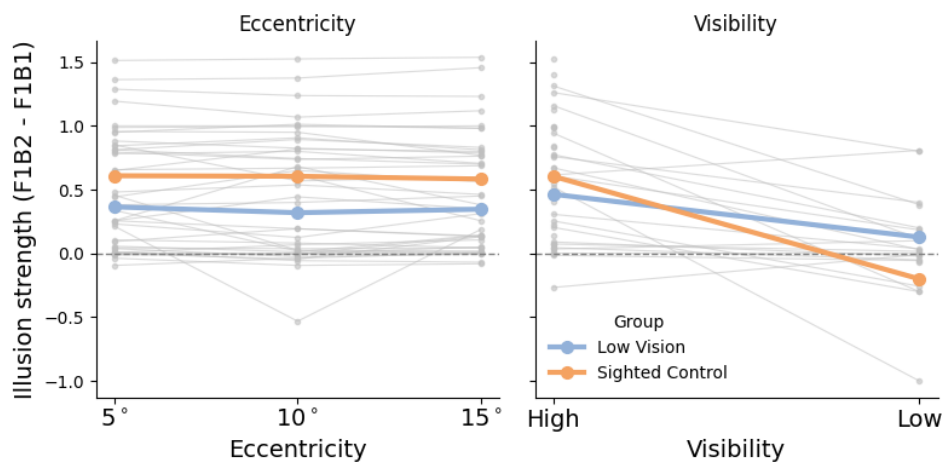


Figure A.5: **Illusion strength as a function of stimulus location.** Illusion strength is plotted across eccentricity (5° – 15°) and visibility levels (high vs. low). No significant effect of eccentricity or visibility was observed based on mixed-effects modeling, although a slight, non-significant decrease in illusion strength is visible under low-visibility conditions.

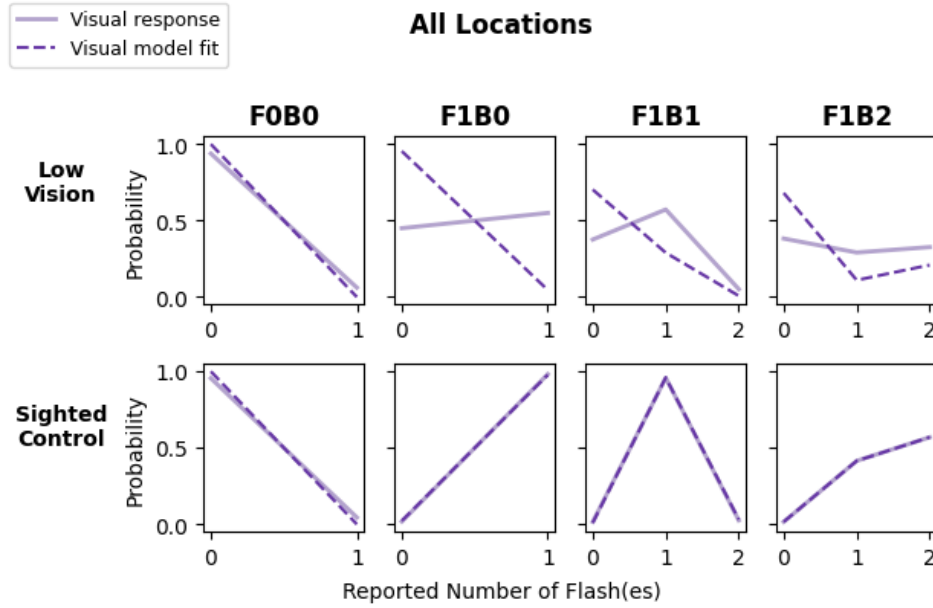


Figure A.6: **Group average Bayesian Causal Inference model fits for all locations.** Visual response distributions (solid lines) and corresponding model-predicted distributions (dashed lines) are shown for four conditions (F0B0, F1B0, F1B1, F1B2). The top row depicts group average for low vision participants, and the bottom row depicts group average for sighted controls. Each panel shows the probability of reporting 0–2 flashes for that condition.

Table A.1: BCI Parameter Estimates and Group Comparisons (All Locations)

Parameter	Sighted Control (Mean $\pm$ SE)	Low Vision (Mean $\pm$ SE)	t	p_{uncorr}	p_{adj}	BF_{10}
p_{common}	0.67 ± 0.08	0.34 ± 0.07	-3.09	< .01	0.01	11.01
μ_P	1.37 ± 0.10	0.73 ± 0.19	-2.98	< .01	0.02	8.74
σ_P	0.79 ± 0.05	1.00 ± 0.17	1.22	0.23	0.92	0.55
σ_V	1.55 ± 0.22	1.26 ± 0.18	-1.01	0.32	1.00	0.46

Note: $N_{Control} = 21$, $N_{LV} = 21$. p_{adj} reflects Bonferroni correction for 4 comparisons. $BF_{10} > 3$ indicates evidence favoring a group difference; $BF_{10} < 1$ indicates evidence favoring the null hypothesis of group invariance.

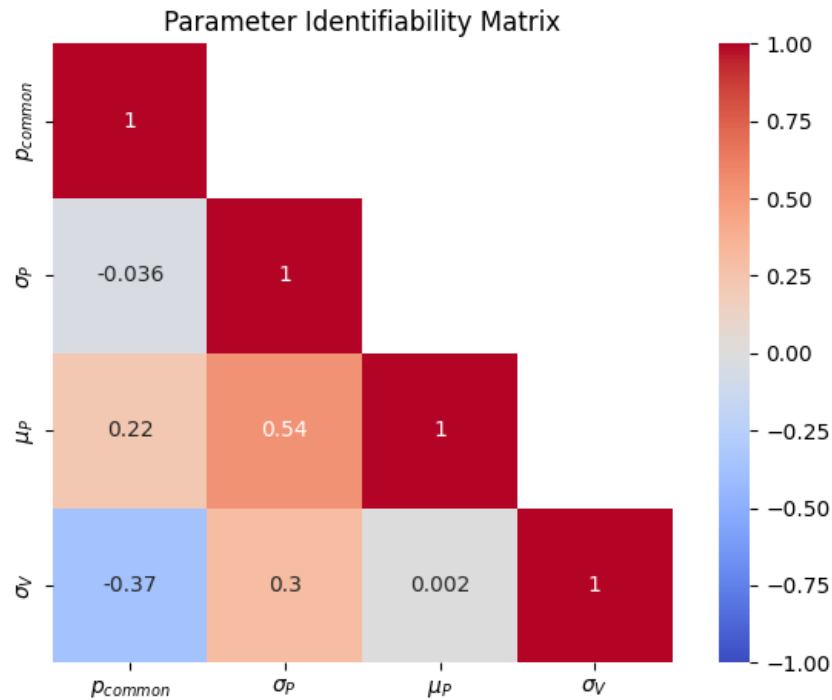


Figure A.7: **Parameter Identifiability Matrix.** Correlation matrix for BCI parameters in high-visibility locations. Most parameter pairs showed weak correlations ($|r| < 0.4$), including the relationship between p_{common} and σ_V ($r = -0.37$), suggesting that visual uncertainty and binding tendency were largely dissociable within the fitting procedure. A moderate correlation was observed between σ_P and μ_P ($r = 0.54$), indicating partial dependence between prior width and prior mean estimates. Overall, the absence of strong inter-parameter correlations supports reasonable parameter identifiability within the BCI framework, particularly for the key distinction between sensory precision and causal binding tendency.

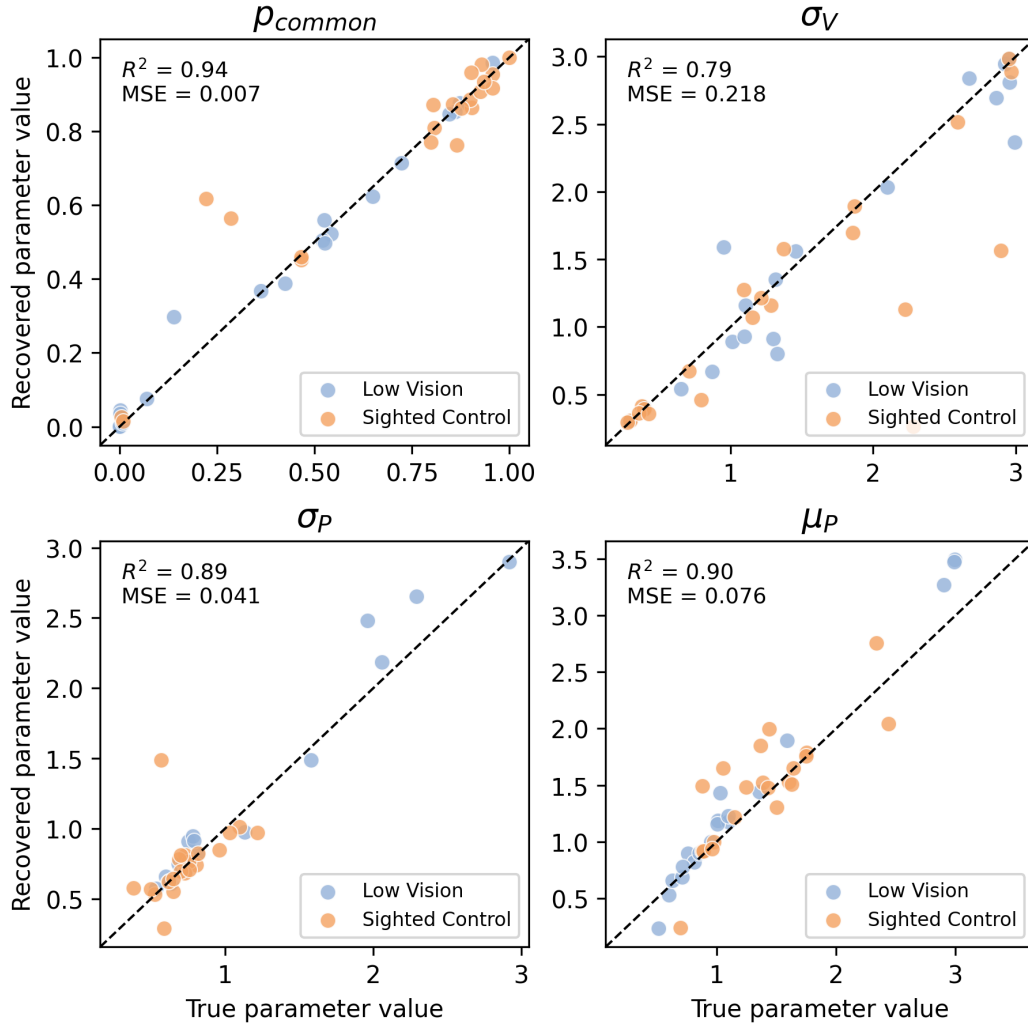


Figure A.8: **Parameter recovery analysis.** Correlation between ground-truth and recovered parameter values for the BCI model. Synthetic datasets were generated using each participant's estimated parameters and subsequently re-fitted to evaluate the model's identifiability and the study design's robustness. In each panel, the recovered values are plotted against the ground truth (black dashed lines indicate the identity $y = x$). High R^2 and low Mean Squared Error (MSE) were observed for all parameters, indicating these parameters were reliably recovered.

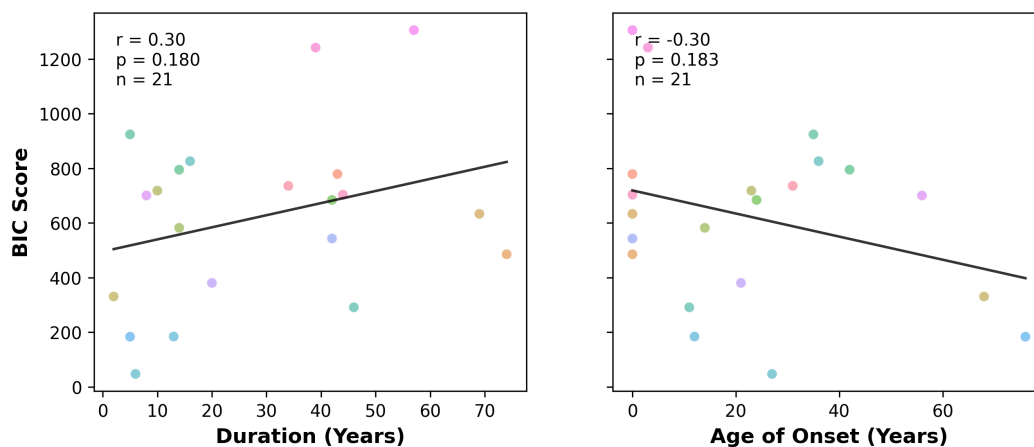


Figure A.9: **Correlation between Disease Characteristics and BIC Scores (Model Performance; the lower the better).** There are no significant correlations between disease duration and BIC score, as well as age of onset and BIC score.

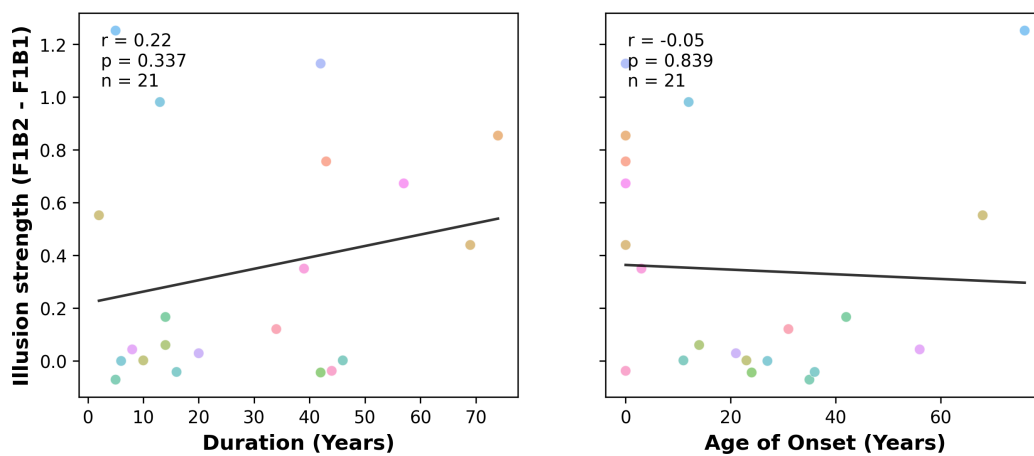


Figure A.10: **Correlation between Disease Characteristics and Illusion Strength.** There are no significant correlations between disease duration and illusion strength (F1B2 - F1B1), as well as age of onset and illusion strength.

Appendix B

SUPPLEMENTAL INFORMATION FOR CHAPTER IV

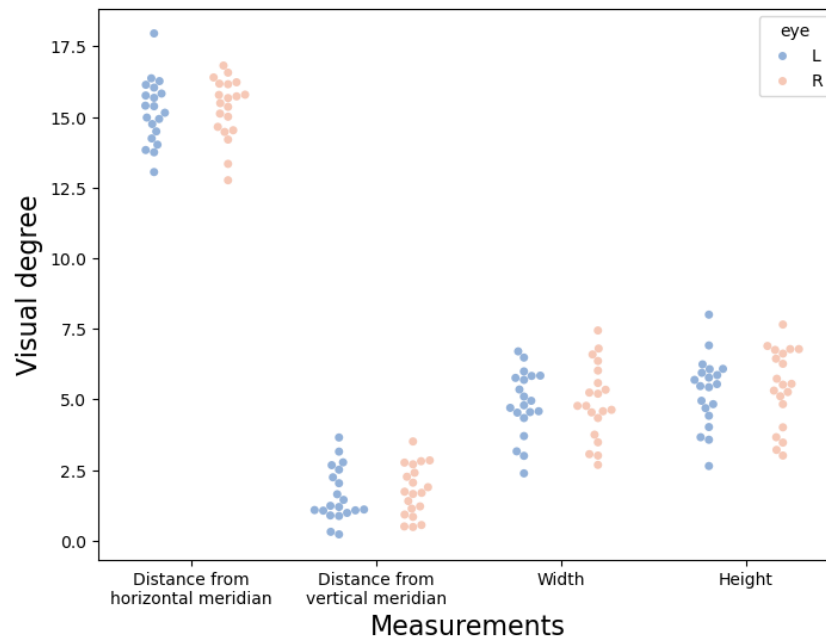


Figure B.1: **Blind Spot Measurements (N = 20)**. The distribution of distances from the vertical and horizontal meridians, as well as the width and height of the blind spots for each eye (left eye: blue, right eye: peach). Here, the vertical and horizontal meridians refer to the midlines of the monitor along the x- and y-axes. Blind-spot locations and sizes were consistent across individuals. The mean blind-spot area was 21.6deg^2 ($SD = 8.99$). Blind-spot centers were located approximately 15.2° from the fovea ($SD = 0.976$) and 1.69° below the horizontal meridian ($SD = 0.780$).

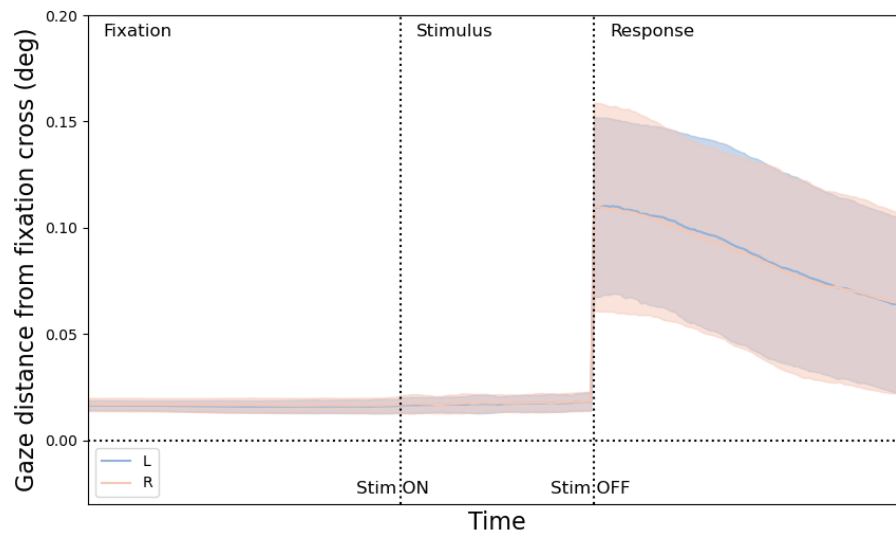


Figure B.2: Eye-tracking measurements during fixation, stimulus presentation, and response phases (N = 20) The y-axis represents the gaze-point distance from the fixation cross (center) in visual degrees, and the x-axis represents time. Solid lines indicate the mean gaze distance across trials, and shaded regions indicate standard deviations. Color code: left eye – blue; right eye – peach. Participants maintained excellent fixation throughout the fixation and stimulus presentation phases. Mean gaze-point distance from the center was 0.016° ($SD = 0.0028$) for the left eye and 0.017° ($SD = 0.0034$) for the right eye during fixation, and 0.017° ($SD = 0.0039$) and 0.018° ($SD = 0.0043$) during stimulus presentation, respectively. Although gaze deviations increased slightly during the response phase (e.g., when reading the text prompt or locating response keys), fixation remained stable during stimulus presentation, confirming that flashes were presented accurately at the intended retinal locations (borders and center of the blind spots). Individual eye-tracking data for all participants are shown in Figure B.3.

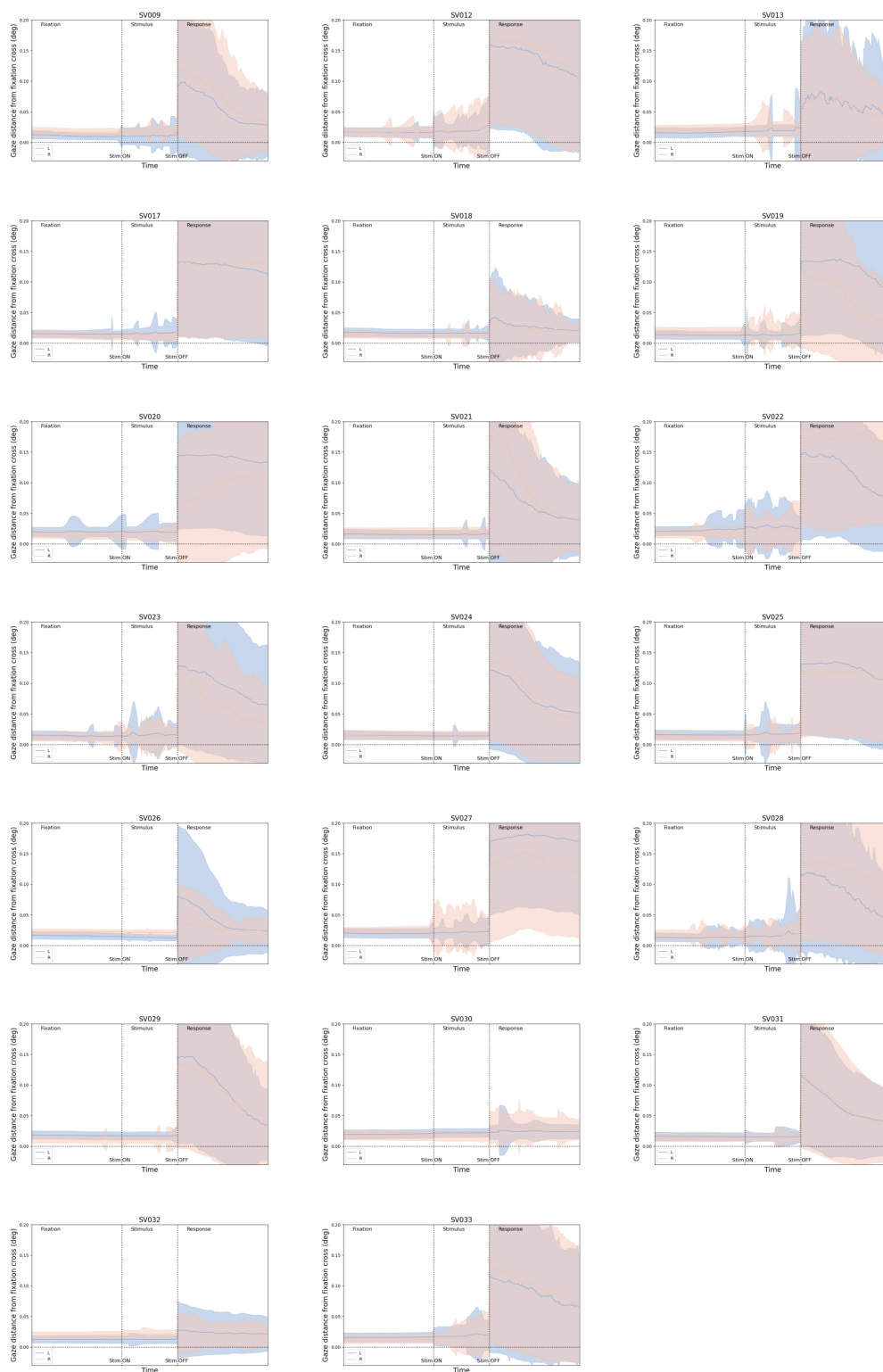


Figure B.3: Individual eye tracking measurements during fixation, stimulus presentation, and response phases.

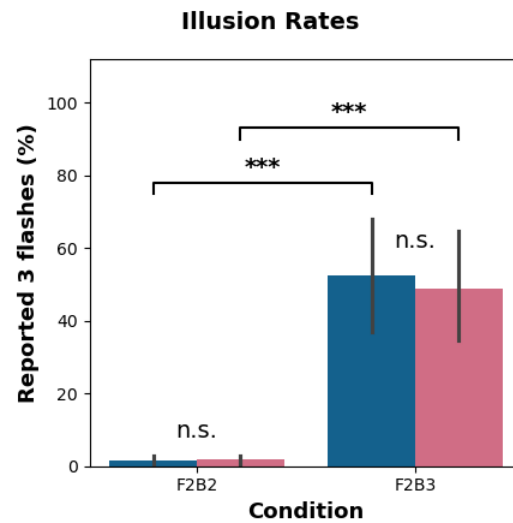


Figure B.4: **Illusion rates in control (F2B2) and illusion (F2B3) conditions.** There is a significant increase in 3-flash responses in the illusion condition compared to the control condition. No significant differences are observed between stimulus locations.

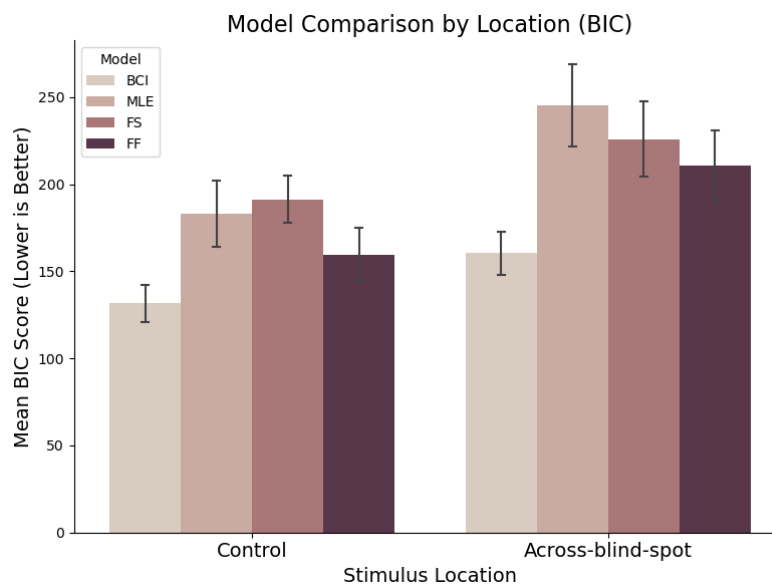


Figure B.5: **Model Comparison by Bayesian Information Criterion (BIC).** Mean BIC scores (lower is better) for the four models tested: BCI, MLE, FS, FF. Error bars represent ± 1 standard error of the mean (SEM).

Table B.1: Comparison between three decision strategies of the BCI model (model averaging, probability matching, and model selection).

Strategy	Location	p_{common}	μ_P	σ_P	σ_V	σ_A	R^2	BIC	% win
Model Averaging	Across-blind-spot	0.55 ± 0.07	1.67 ± 0.17	1.12 ± 0.15	1.65 ± 0.19	0.25 ± 0.03	0.96 ± 0.02	184.43 ± 17.41	45.0
	Control	0.59 ± 0.08	2.28 ± 0.17	0.87 ± 0.08	1.12 ± 0.14	0.33 ± 0.13	0.95 ± 0.03	146.64 ± 15.38	45.0
Probability Matching	Across-blind-spot	0.49 ± 0.08	1.76 ± 0.16	1.09 ± 0.12	1.79 ± 0.20	0.20 ± 0.02	0.97 ± 0.01	167.94 ± 12.19	20.0
	Control	0.59 ± 0.08	2.26 ± 0.18	0.82 ± 0.07	1.12 ± 0.15	0.19 ± 0.01	0.98 ± 0.00	135.07 ± 10.33	30.0
Model Selection	Across-blind-spot	0.49 ± 0.05	1.91 ± 0.14	1.05 ± 0.14	1.92 ± 0.16	0.23 ± 0.02	0.97 ± 0.01	166.67 ± 13.16	35.0
	Control	0.60 ± 0.06	2.33 ± 0.16	0.74 ± 0.08	1.05 ± 0.11	0.20 ± 0.01	0.98 ± 0.00	133.99 ± 11.09	25.0

Note: p_{common} , causal prior; μ_P , mean of the numeric prior; σ_P , standard deviation of the numeric prior; σ_V , standard deviation of the visual likelihood; σ_A , standard deviation of the auditory likelihood; R^2 , coefficient of determination; % win, percentage of participants in which a model won the within-participant model comparison based on BIC.

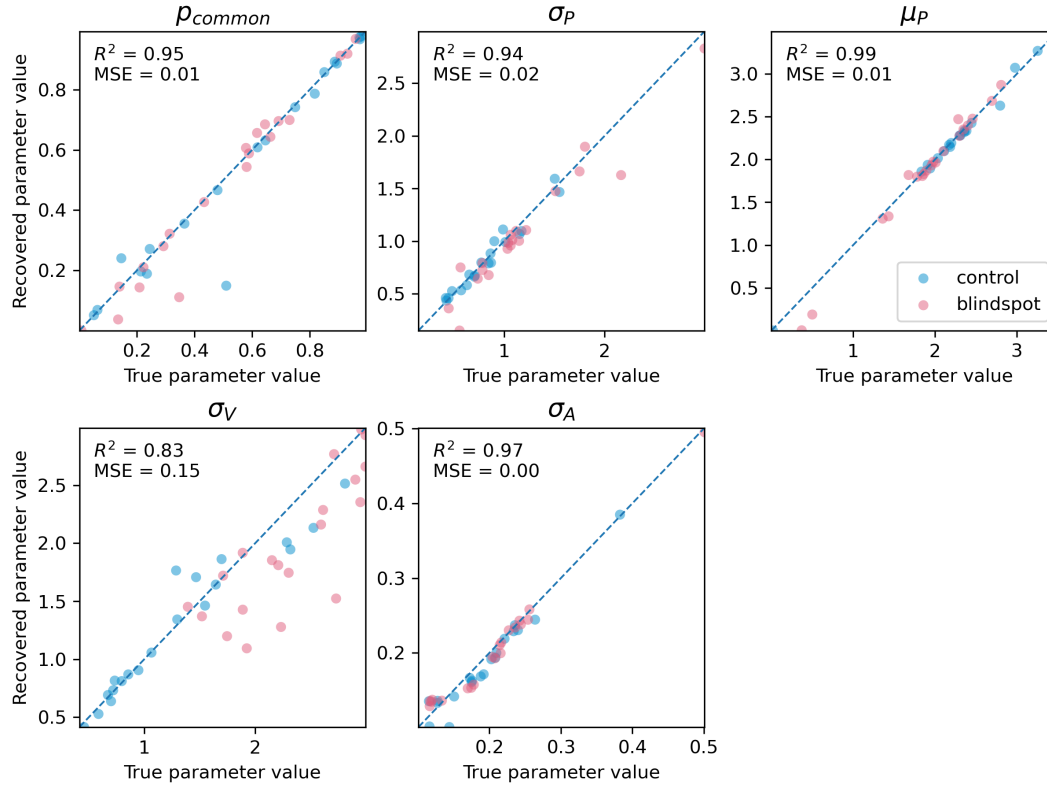


Figure B.6: **Parameter recovery analysis.** Correlation between ground-truth and recovered parameter values for the BCI model. Synthetic datasets were generated using each participant's estimated parameters and subsequently re-fitted to evaluate the model's identifiability and the study design's robustness. In each panel, the recovered values are plotted against the ground truth (black dashed lines indicate the identity $y = x$). High R^2 and low Mean Squared Error (MSE) across all panels demonstrate that the model parameters were reliably recovered.

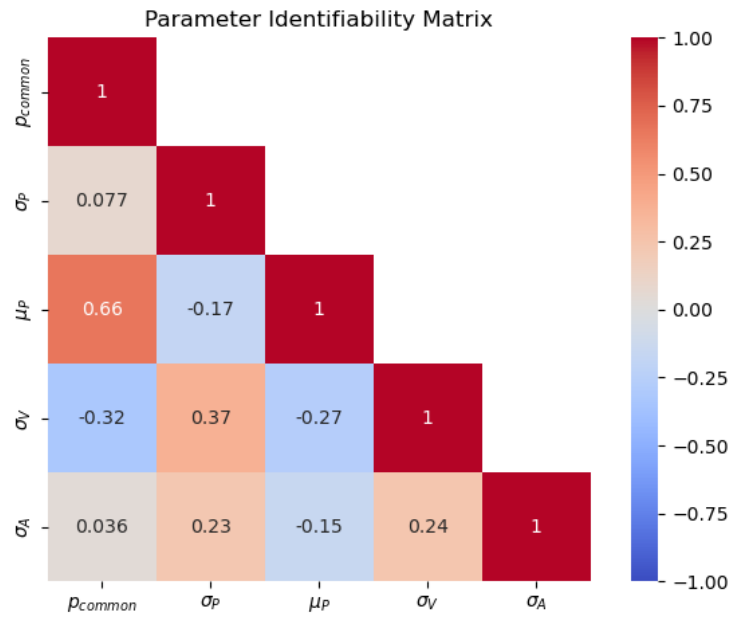


Figure B.7: **Parameter Identifiability Matrix.** Pearson correlation coefficients (r) between best-fitting BCI model parameters pooled across participants and locations ($N = 40$). Visual noise (σ_V), numerical prior width (σ_P), and auditory noise (σ_A) show minimal inter-parameter correlation, indicating high identifiability. A significant correlation cluster was observed between the causal prior (P_{common}) and numerical prior mean (μ_P), reflecting a functional coupling between causal belief and numerical expectation in the absence of direct visual input.

Appendix C

SUPPLEMENTAL INFORMATION FOR CHAPTER V

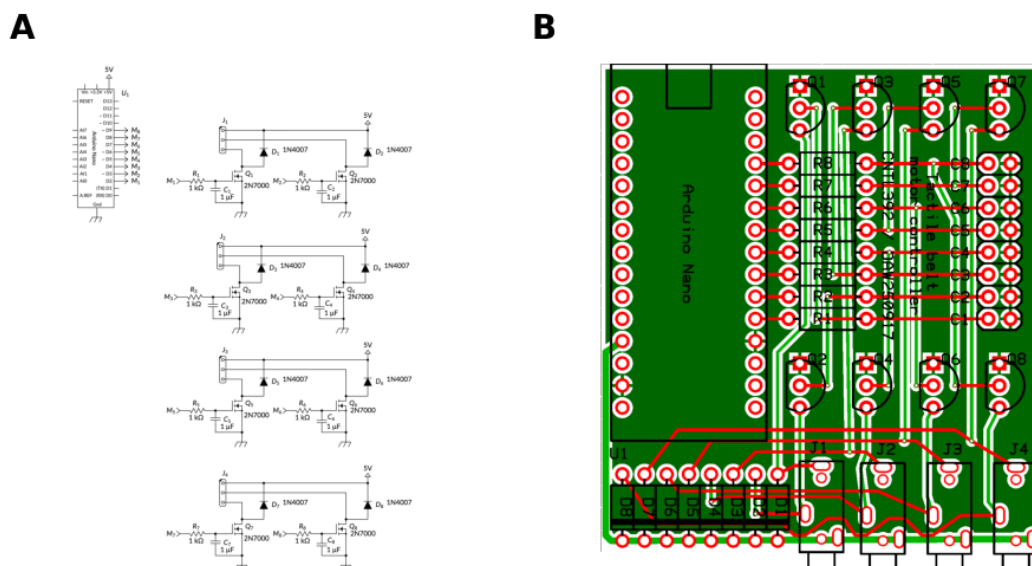
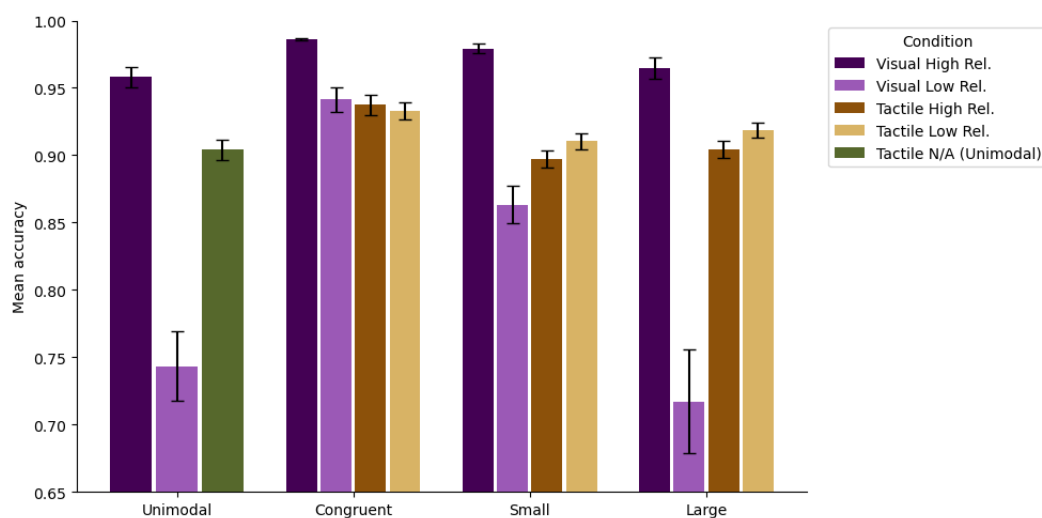


Figure C.1: Detailed circuit schematics and PCB layout.

Figure C.2: Mean accuracy across conditions, all directions pooled. Error bars represent ± 1 SEM.

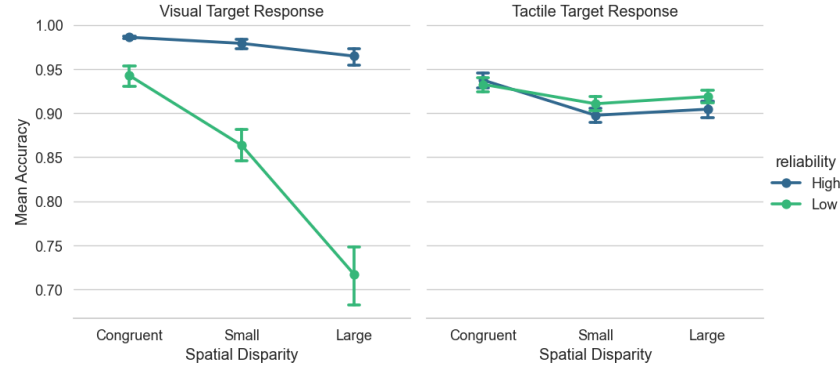


Figure C.3: Impact of Spatial Disparity and Visual Reliability on Accuracy. Mean localization accuracy is shown for the (Left) Visual Target and (Right) Tactile Target tasks. Accuracy is normalized to a $[0, 1]$ scale ($1.0 = 0^\circ$ error); error bars represent ± 1 SEM. A robust three-way interaction was observed ($p < .001$) for visual response, characterized by a dramatic decline in accuracy as spatial disparity increased under low visual reliability (light green). On the other hand, accuracy remained relatively stable and high across visual and spatial manipulations for tactile response.

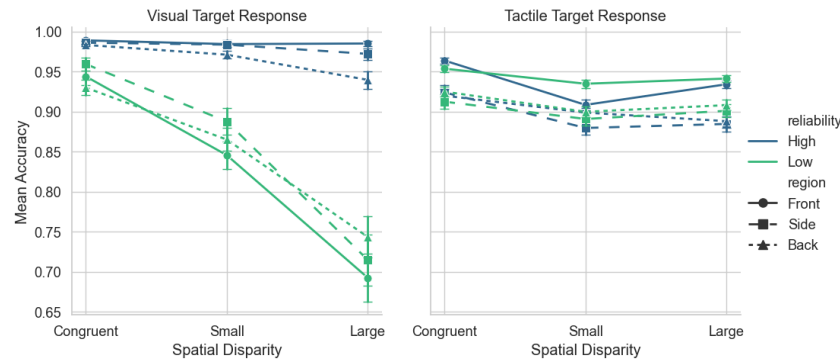


Figure C.4: Impact of Spatial Disparity, Visual Reliability, and Region on Accuracy. Mean localization accuracy is shown for the (Left) Visual Target and (Right) Tactile Target tasks. Accuracy is normalized to a $[0, 1]$ scale ($1.0 = 0^\circ$ error); error bars represent ± 1 SEM. We observe similar trends from Fig. C.3. There is no significant 4-way interaction, indicating that these performance trends were consistent across all spatial Regions (Front, Side, and Back).

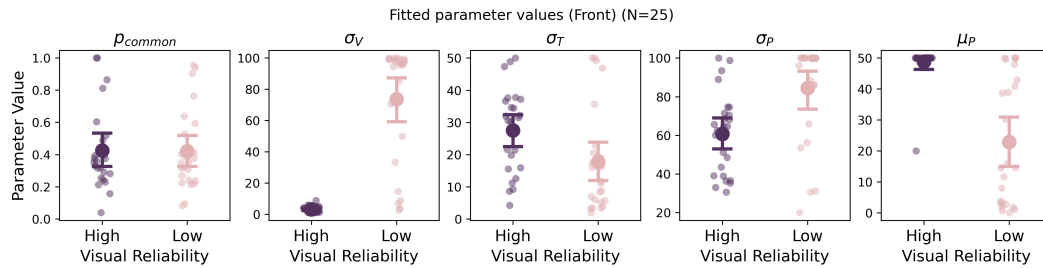


Figure C.5: **Fitted BCI Parameters.** Parameter values evaluated separately across the Front region under High and Low visual reliability. Large markers represent the group mean (± 1 SEM), with individual participant data overlaid.

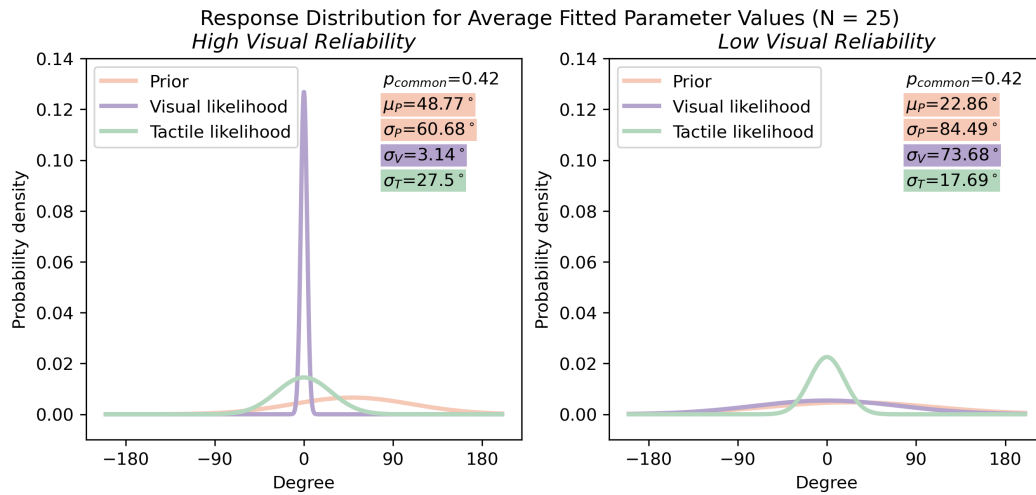


Figure C.6: **Response Distributions.** Probability density curves mapping the visual likelihood, tactile likelihood, and spatial prior across the Front region. Consistent with the global fits, the spatial priors remain notably flat across all spatial areas, negating the behavioral relevance of shifts in μ_P .

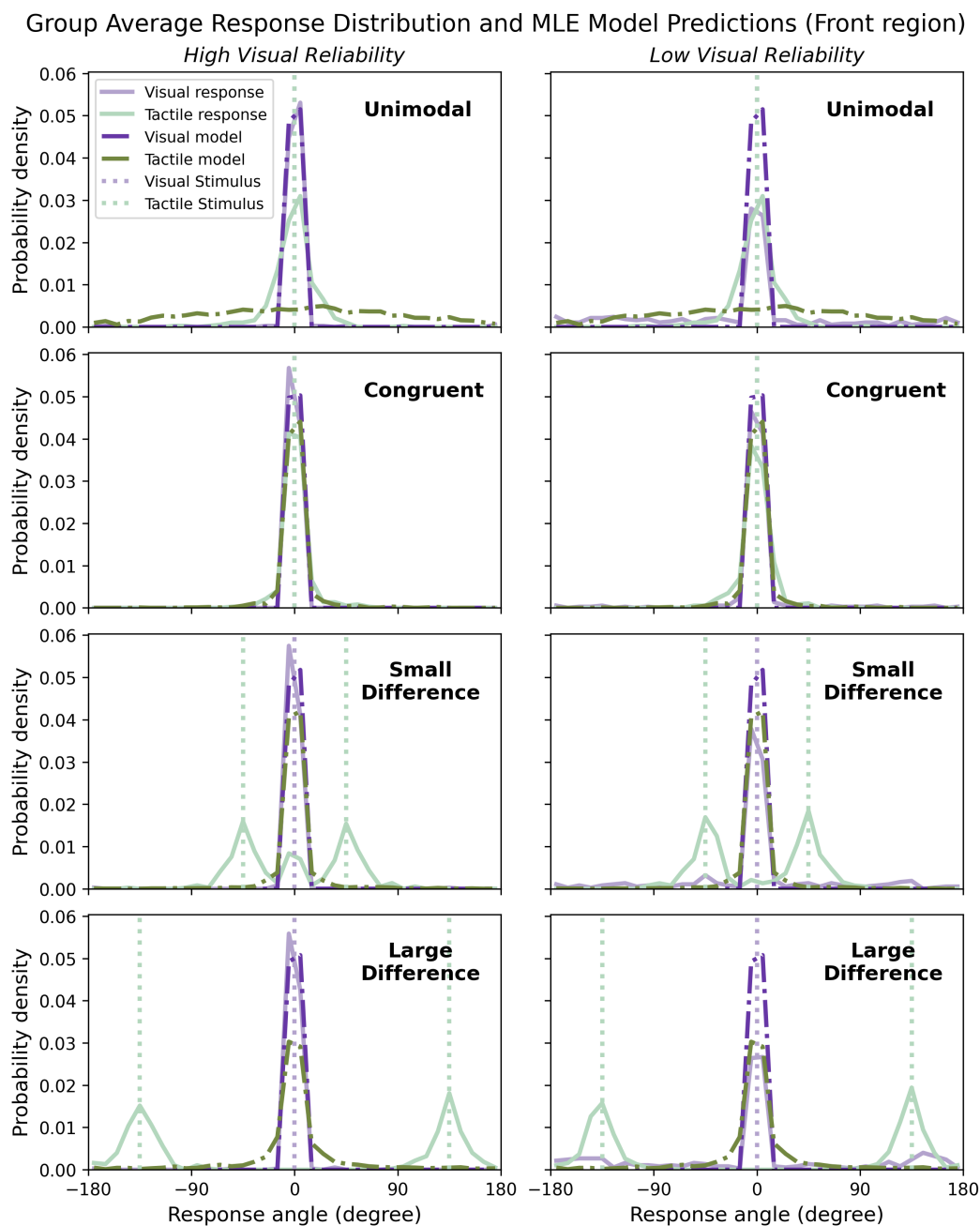


Figure C.7: **Group average MLE model fits.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0°.

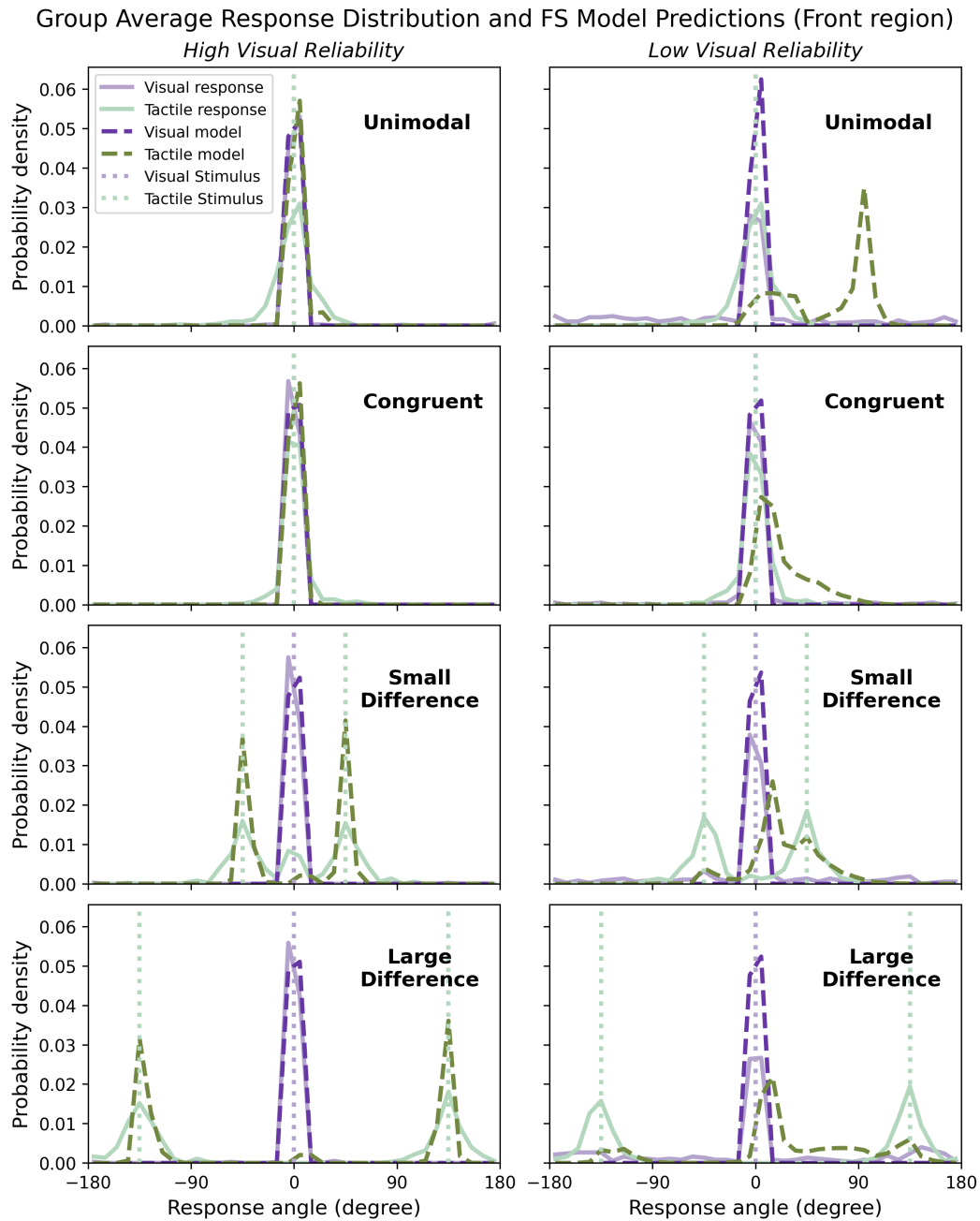


Figure C.8: **Group average FS model fits.** Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° .

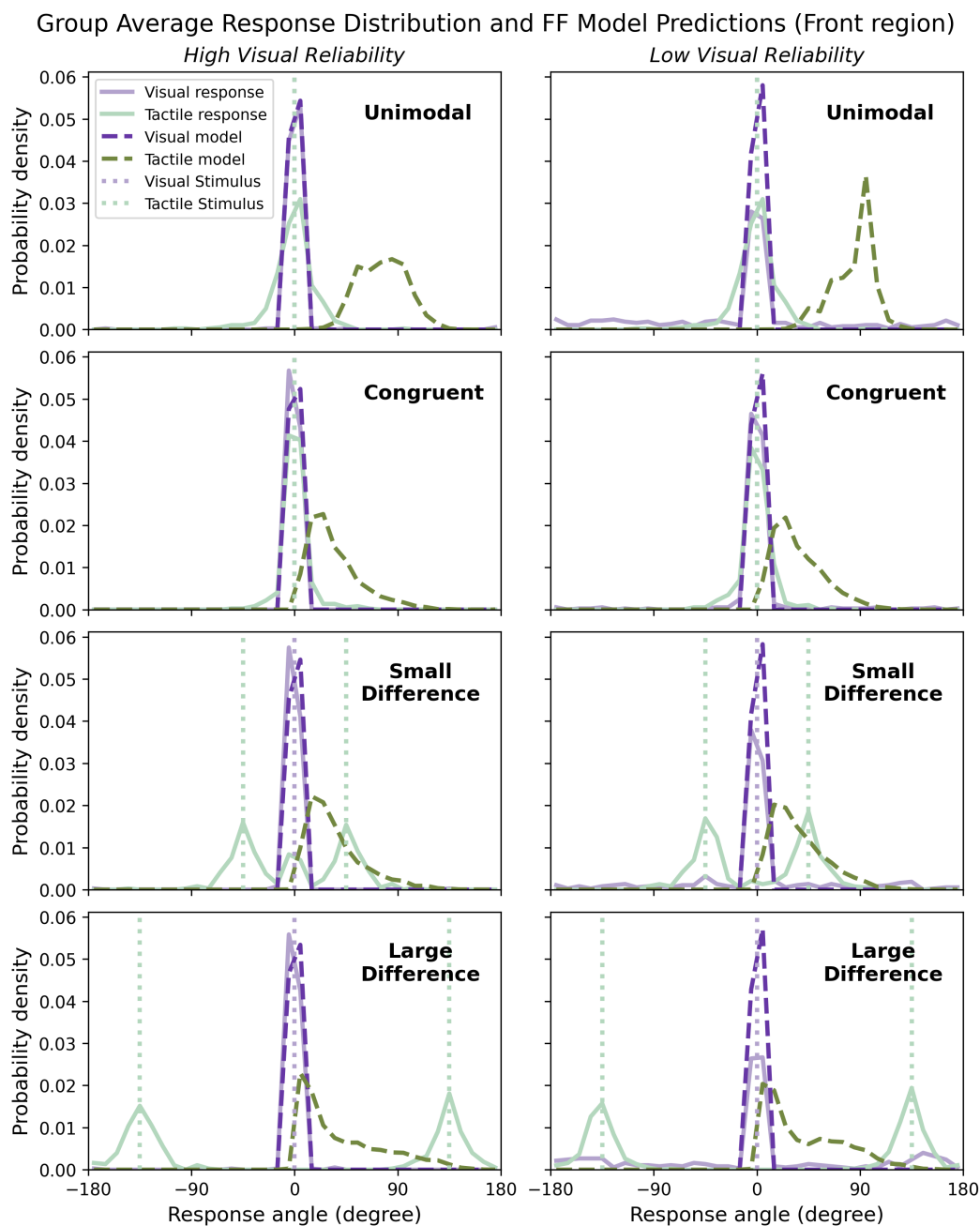


Figure C.9: Group average FF model fits. Solid purple and green lines correspond to the visual and tactile response data, dashed purple and green lines correspond to the model predictions for visual and auditory responses, and dotted purple and green lines correspond to true stimulus values. All true and stimulated response data are centered and normalized at 0° .

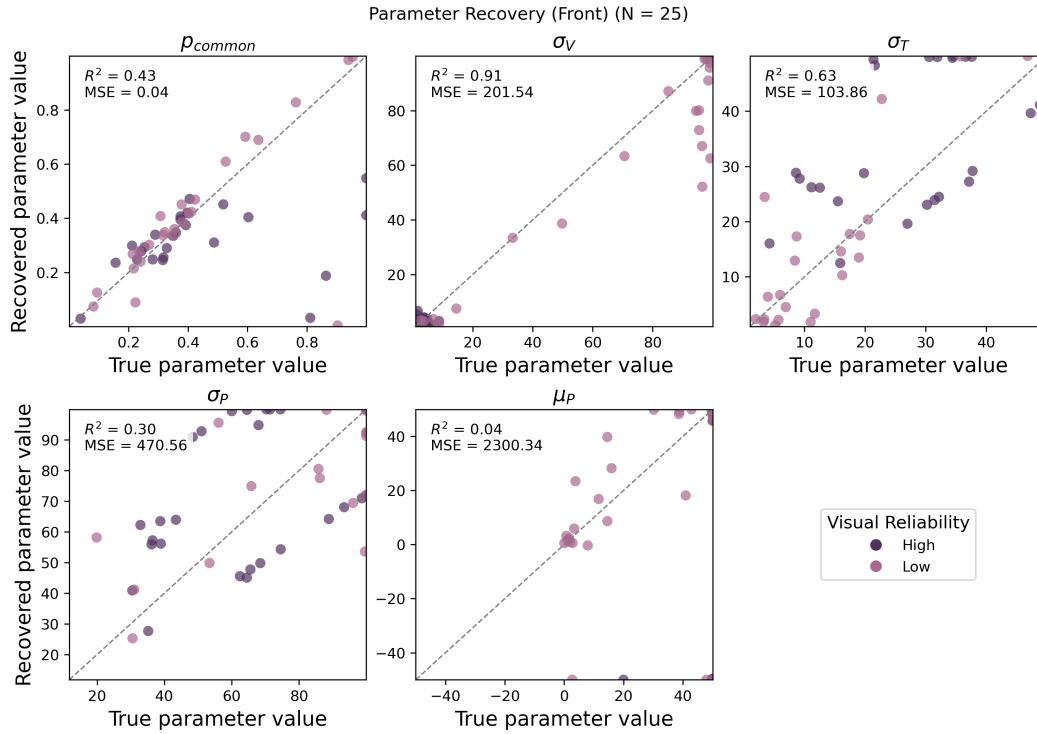


Figure C.10: Parameter recovery analysis for the Bayesian Causal Inference (BCI) model. Scatter plots illustrate the correspondence between ground-truth and recovered parameter values across 25 synthetic datasets. To evaluate model identifiability and experimental robustness, synthetic data were generated using each participant's estimated parameters and subsequently re-fitted. In each panel, recovered values are plotted against ground truth, with dashed gray lines representing the identity ($y = x$). Sensory likelihood parameters (σ_V and σ_T) show strong recovery, with σ_V achieving an $R^2 = 0.91$. While p_{common} displays a lower R^2 (0.43), its remarkably low Mean Squared Error (MSE = 0.04) indicates that estimates are highly accurate but clustered within a narrow range, making R^2 sensitive to minor deviations. The relatively poor recovery of spatial prior parameters (σ_P, μ_P)—characterized by low R^2 and high MSE—is expected, as the fitted priors were sufficiently large to be uninformative, rendering the model less sensitive to their specific values.

Parameter Identifiability Matrix (Front)

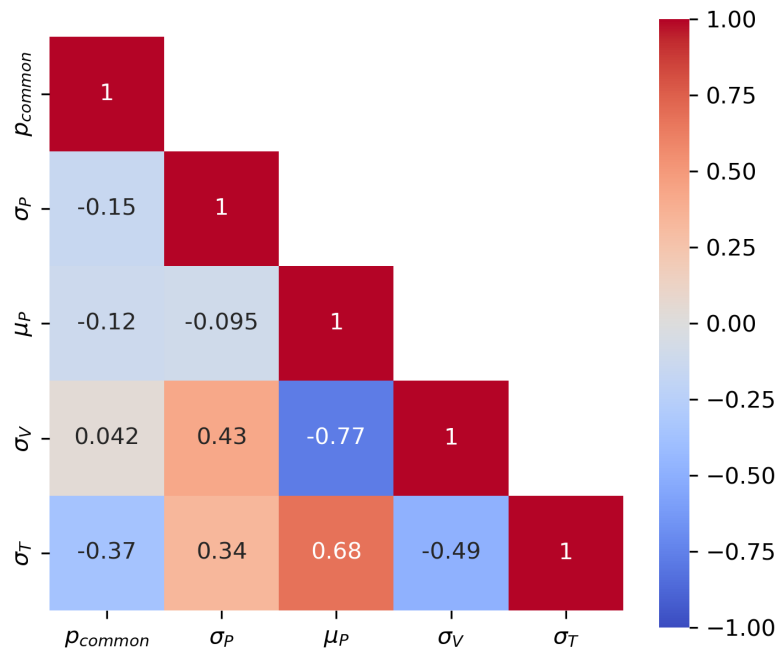


Figure C.11: **Parameter Identifiability Matrices.** Pearson correlation coefficients (r) calculated for the Front region ($N = 25$). The primary parameters (p_{common} , σ_P , σ_V) exhibit minimal inter-parameter correlation, confirming that the BCI framework uniquely identifies these values regardless of spatial location. Observed correlations between visual noise (σ_V) or tactile noise (σ_T) and the spatial prior mean (μ_P) are considered functionally negligible. Given the uninformatively broad spatial prior (σ_P) and the limited recoverability of μ_P , these spatial offsets do not meaningfully influence the resulting perceptual estimates, nor do they compromise the stability of the core causal and sensory parameters.

