Sensory reliability shapes perceptual inference via two mechanisms

Tim Rohe
Max Planck Institute for Biological Cybernetics,
Tuebingen, Germany

Uta Noppeney
Max Planck Institute for Biological Cybernetics,
Tuebingen, Germany
Computational Neuroscience and Cognitive Robotics Centre,
University of Birmingham, Birmingham, UK

To obtain a coherent percept of the environment, the brain should integrate sensory signals from common sources and segregate those from independent sources. Recent research has demonstrated that humans integrate audiovisual information during spatial localization consistent with Bayesian Causal Inference (CI). However, the decision strategies that human observers employ for implicit and explicit CI remain unclear. Further, despite the key role of sensory reliability in multisensory integration, Bayesian CI has never been evaluated across a wide range of sensory reliabilities. This psychophysics study presented participants with spatially congruent and discrepant audiovisual signals at four levels of visual reliability. Participants localized the auditory signals (implicit CI) and judged whether auditory and visual signals came from common or independent sources (explicit CI). Our results demonstrate that humans employ model averaging as a decision strategy for implicit CI; they report an auditory spatial estimate that averages the spatial estimates under the two causal structures weighted by their posterior probabilities. Likewise, they explicitly infer a common source during the common-source judgment when the posterior probability for a common source exceeds a fixed threshold of 0.5. Critically, sensory reliability shapes multisensory integration in Bayesian CI via two distinct mechanisms: First, higher sensory reliability sensitizes humans to spatial disparity and thereby sharpens their multisensory integration window. Second, sensory reliability determines the relative signal weights in multisensory integration under the assumption of a common source. In conclusion, our results demonstrate that Bayesian CI is fundamental for integrating signals of variable reliabilities.

Introduction

Imagine you are engaged in a conversation at a busy party. You will understand your conversational partner more clearly when you integrate the acoustic speech with his facial articulatory movements. By contrast, speech comprehension will deteriorate if you erroneously integrate his facial movements with another person's acoustic speech signal. Thus, audiovisual integration requires the brain to infer whether signals come from common or independent sources. This challenge cannot be addressed by traditional forced-fusion models that enforce signals to be integrated in a mandatory fashion (Ernst & Banks, 2002) but requires Bayesian Causal Inference (CI) that explicitly models the potential causal structures that could have generated the sensory signals (Koerding et al., 2007; Shams & Beierholm, 2010). In the case of a common source, the sensory signals are integrated weighted by their reliability into the most reliable unbiased estimate. In the case of separate sources, signals are processed independently. Importantly, the brain does not know the underlying causal structure, but needs to infer it from the sensory signals based on spatial, temporal, and structural correspondences (Gepshtein, Burge, Ernst, & Banks, 2005; Lewald & Guski, 2003; Slutsky & Recanzone, 2001; Wallace et al., 2004). A final estimate of a physical property is obtained by combining the estimates under the various causal structures using decisional strategies such as model averaging, model selection, or probability matching (Wozny, Beierholm, & Shams, 2010).

Previous modeling efforts have demonstrated that humans integrate information for spatial localization consistent with Bayesian CI (Koerding et al., 2007;

Citation: Rohe, T., & Noppeney, U. (2015). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, *15*(5):22, 1–16, http://www.journalofvision.org/content/15/5/22, doi:10.1167/15.5.22.

doi: 10.1167/15.5.22 Received March 26, 2014; published April 29, 2015 ISSN 1534-7362 © 2015 ARVO

Wozny et al., 2010). For small spatial discrepancies, the perceived location of an auditory event shifts towards the location of a temporally correlated but spatially displaced visual event and vice versa, depending on the relative auditory and visual reliabilities (Alais & Burr, 2004). Yet, for large spatial discrepancies, when it is unlikely that audiovisual signals arise from a common source, these crossmodal biases are greatly attenuated (Wallace et al., 2004). Moreover, when participants indicated that the audiovisual signals came from independent sources, the perceived auditory location shifted less towards or even away from the true visual location (Koerding et al., 2007; Wallace et al., 2004).

However, so far Bayesian CI models have been applied to psychophysics data that included only one or two reliability levels (Beierholm, Quartz, & Shams, 2009). Given the key role of reliability in multisensory integration, it is critical to demonstrate that Bayesian CI predicts observers' response profile when sensory signals vary in their reliability over a wide range. Furthermore, it is unclear how participants perform CI decisions implicitly during spatial localization and explicitly during common-source judgments. For audiovisual spatial localization, one recent study has suggested that humans do not perform model averaging as previously assumed, but employ a suboptimal strategy of probability matching (Wozny et al., 2010). In other words, they report the spatial estimate of one particular causal structure sampled in proportion to the posterior probability of this causal structure.

Yet, it is unclear whether a similar decision strategy is employed when CI decisions are invoked explicitly in common-source judgments. As implicit and explicit CI tasks serve different goals, they may be governed by different utility functions associated with different decision strategies. It is conceivable that implicit and explicit CI access the same posterior common-source probability, yet use it with different decision strategies.

To address these questions, we presented participants with spatially congruent and discrepant audiovisual signals at four visual reliability levels in a spatial ventriloquist paradigm. On each trial, participants located the auditory signal and judged whether the audiovisual signals emanated from a common source. We then fitted the Bayesian CI model commonly to spatial localization and common-source judgments under various decision strategies.

Methods

Subjects

Twenty-six healthy subjects participated in the study after giving informed consent (16 female, $M =$ 25.8 years, range 23–37 years). All subjects had normal or corrected-to-normal vision and reported normal hearing. The study was approved by the ethics committee of the University of Tübingen (protocol number 432 2007 BO1) and adhered to the Declaration of Helsinki.

Stimuli

The visual stimulus was a cloud of 20 white dots (diameter: 0.43° visual angle; luminance 91 cd/m^2) sampled from a bivariate Gaussian presented on a dark gray background (luminance 62 cd/m^2; i.e., 47% contrast). The vertical standard deviation of the Gaussian was set to 5.4°. To manipulate the spatial reliability of the visual signal, the horizontal standard deviation was set to four levels: 0.1°, 5.4°, 10.8°, or 16.2°. A separate experiment with a different set of subjects demonstrated that the physical standard deviation correlated with the perceptual visual standard deviation estimated from a purely unisensory visual localization task (see Supplementary Experiment 1 and Table S4). To manipulate the spatial location of the visual stimulus, the mean of the Gaussian was sampled from five possible locations along the azimuth (i.e., −10°, −5°, 0°, 5°, or 10°). The auditory spatial signal was a burst of white noise. To create a virtual auditory spatial signal, the noise was convolved with spatially specific head-related transfer functions (HRTFs). The HRTFs were pseudo-individualized by matching subjects' head width, height, and depth to the anthropometry of subjects in the Center for Image Processing and Integrated Computing (CIPIC) database (Algazi, Duda, Thompson, & Avendano, 2001). HRTFs from the available locations in the database were interpolated to the desired locations of the auditory signal. We used pseudo-individualized HRTFs that have been shown to enable localization accuracies that are comparable to those of speaker-based free-field sounds (Wenzel, Arruda, Kistler, & Wightman, 1993) and of signals generated from individual HRTFs (Wightman & Kistler, 1989). Furthermore, we confirmed that participants obtained high sound localization accuracy on 25 unisensory auditory practice trials prior to the main experiment (across-subjects' mean correlation coefficient between perceived and true sound location: 0.829).

Experimental design and procedure

In a spatial ventriloquist paradigm (Figure 1A), participants were presented with synchronous, yet spatially congruent or discrepant visual and auditory signals. On each trial, auditory and visual locations were independently sampled from five possible locations along the azimuth (i.e., –10°, –5°, 0°, 5°, or 10°). In addition, we manipulated the reliability of the visual signal by setting the horizontal standard deviation of the Gaussian cloud to one of four possible levels (i.e., 0.1°, 5.4°, 10.8°, or 16.2° *SD*). Hence, our experiment included 100 conditions arranged in a 5 (auditory location: S_A) × 5 (visual location: S_V) × 4 (visual reliability: $1/\sigma_V{}^2$) factorial design.

On each trial, synchronous auditory and visual spatial signals were presented for 33 ms. Participants responded to two questions presented sequentially: First, participants localized the auditory spatial signal as accurately as possible by pushing one of five buttons that corresponded spatially to the stimulus locations (i.e., spatial localization). Second, participants decided whether the visual and auditory signals were generated by common or independent sources (i.e., common-source judgment) and indicated their response via a two-choice key press. Participants were instructed to use the perceived center of the Gaussian cloud as reference for their judgments. The time limit for both responses was 3 s. The next trial started with a variable interval of 1–2.5 s after participants had given their second button response. Throughout the experiment, participants fixated a cross (1.5° diameter) presented in the center of the screen.

The locations of the auditory and visual signals were independently randomized across trials. Each combination of auditory and visual locations was presented equally often in each participant. The levels of visual reliability were presented either in blocks (55–85 trials per level of visual reliability) or varied according to a Markov chain (with a transition probability of 0.1 to change to an adjacent level and of 0.9 to stay at the same level of visual reliability). As the type of sequence did not influence the effects reported in this manuscript, we pooled over the two sequences and analyzed them together.

In total, each participant completed 440–850 experimental trials. Twelve subjects participated in a longer version of the experiment including an additional level of visual reliability (21.6° *SD*). As shown in the supplementary materials, the response profile for this

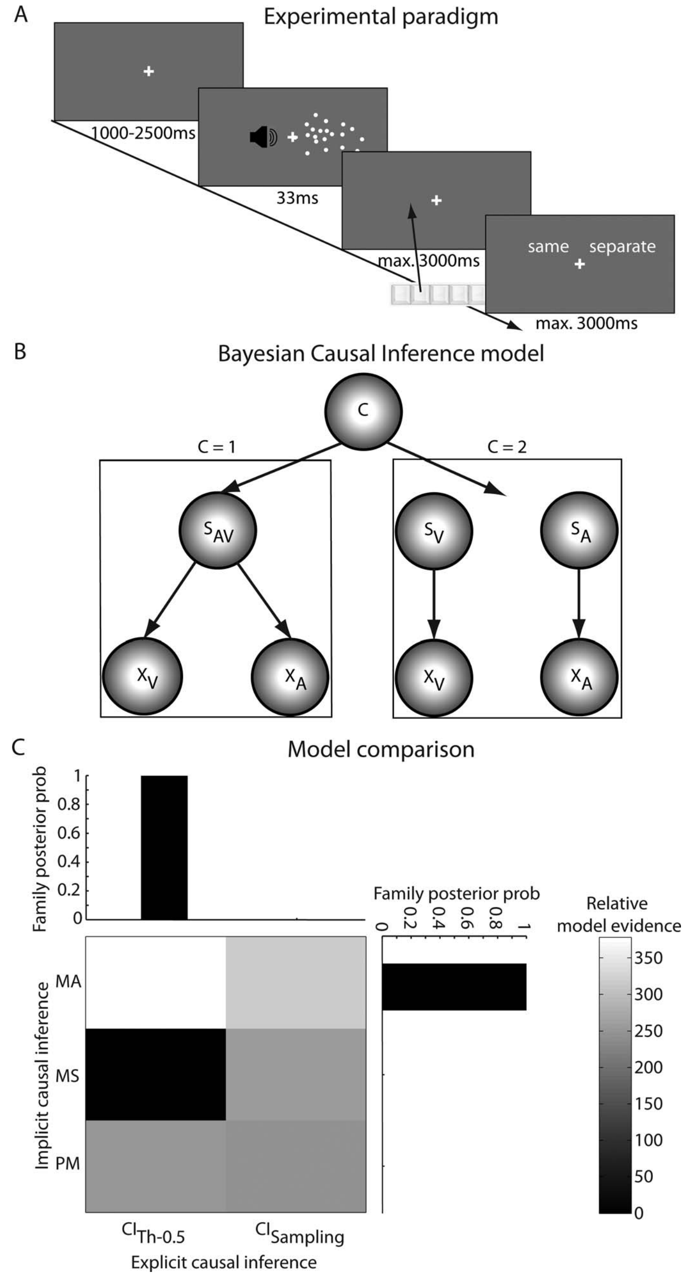

Figure 1. Experimental design, Bayesian CI model and results of the model comparison. (A) Stimuli and time course of an experimental trial in the ventriloquist paradigm. After a variable fixation interval, participants were presented with synchronous, spatially congruent, or discrepant visual and auditory signals along the azimuth. Using five response buttons, participants localized the auditory signal and decided whether the visual and auditory signals were generated by common (same) or independent (separate) sources. (B) In the Bayesian CI model (adapted from Koerding et al., 2007), auditory (x_A) and visual (x_V) spatial signals are generated either by a common ($C = 1$; S_{AV}) or independent ($C = 2$) auditory (S_A) and visual (S_V) sources. (C) The 3 × 2 factorial model space manipulated (i) the implicit CI strategy involved in auditory spatial localization: model averaging (MA), model selection (MS), or probability matching →

← (PM) and (ii) the explicit CI strategy involved in the common-source judgment: a fixed threshold of 0.5 ($\mathrm{CI}_{\mathrm{Th}\text{-}0.5}$) or a sampling strategy ($\mathrm{CI}_{\mathrm{Sampling}}$). The matrix shows the relative group BIC as an approximation to the log model evidence (i.e., across-subjects sum of the individual BICs) of the six models relative to the worst model (larger = better). The bar plots show the family posterior probabilities of the three implicit CI model families (right) and the two explicit CI model families (top). Note that we fitted the six CI models individually to the auditory localization and common-source judgments responses of each participant.

lowest reliability level was consistent with the results for the four levels of visual reliability that are presented in the main paper (see Supplementary Figure S1). However, to obtain comparable data sets from all 26 participants, the data for the lowest reliability level were excluded in the analyses of the main paper. After exclusion of the data from this condition and exclusion of trials with a missed response, 390–720 trials per subject were included in the study. Prior to the main experiment, participants practiced the auditory localization task on 25 unisensory auditory trials, 25 audiovisual congruent trials with a single dot as the visual spatial signal, and 15 trials with stimuli as in the main experiment.

Experimental setup

Audiovisual stimuli were presented using Psychtoolbox 3.09 (Brainard, 1997; Kleiner et al., 2007; www.psychtoolbox.org) running under Matlab R2010b (MathWorks) on a Windows machine (Microsoft XP 2002 SP2). Auditory stimuli were presented at ~75 dB SPL using headphones (HD 555, Sennheiser, Wedemark-Wennebostel, Germany). Because visual stimuli required a large field of view, they were presented on a 30-in. LCD display (UltraSharp 3007WFP, Dell, Round Rock, TX). Participants were seated at a table in front of the screen in a darkened booth, resting their head on an adjustable chin rest. The viewing distance was 27 cm, resulting in a visual field of approximately 100°. Subjects indicated their responses using a standard keyboard. Subjects used the buttons {1, 2, 3, 4, r} for spatial localization responses with their left hand and {9, 0} for common-source judgments with their right hand.

CI model

We employed a CI model of audiovisual perception (Koerding et al., 2007). On each trial, participants performed two tasks, an auditory localization and a common-source judgment. For each of the two tasks, we augmented the CI model with several decision strategies. For the implicit CI involved in the auditory localization task, we employed (a) model averaging, (b) model selection, and (c) probability matching as previously described by Wozny et al. (2010). For the explicit CI involved in the common-source judgment, we used two decision functions that are described in detail below. By manipulating the decision functions for the spatial localization and the common-source judgment in a factorial fashion, we generated a 3 × 2 model space.

Details of the Bayesian generative model can be found in Koerding et al. (2007). Briefly, we assume that a common ($C = 1$) or independent ($C = 2$) source is determined by sampling from a binomial distribution with the common-source prior $P(C = 1) = p_{\text{common}}$ (Figure 1B). For a common source, the "true" location S_{AV} is drawn from the spatial prior distribution $N(\mu_P, \sigma_P)$. For two independent sources, the true auditory (S_A) and visual (S_V) locations are drawn independently from this spatial prior distribution. For the spatial prior distribution, we assumed a central bias (i.e., $\mu_P = 0°$). We introduced sensory noise by independently drawing x_A and x_V from normal distributions centered on the true auditory (and respectively, visual) locations with parameters ${\sigma_A}^2$ (respectively, ${\sigma_V}^2$). Thus, the generative model included the following free parameters: the common-source prior p_{common}, the spatial prior variance ${\sigma_P}^2$, the auditory variance ${\sigma_A}^2$, and the four visual variances ${\sigma_V}^2$ corresponding to the four visual reliability levels.

The probability of the underlying causal structure can be inferred by combining the common-source prior with the sensory evidence according to Bayes rule:

$$p(C = 1|\ x_A, x_V) = \frac{p(x_A, x_V|C = 1)p_{\text{common}}}{p(x_A, x_V)} \tag{1}$$

In the case of a common source ($C = 1$; Figure 1B left), the maximum a posteriori probability estimate of the auditory location is a reliability-weighted average of the auditory and visual estimates and the prior.

$$\hat{S}_{AV,C=1} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}} \tag{2}$$

In the case of a separate-source inference ($C = 2$; Figure 1B right), the estimate of the auditory signal location is independent from the visual spatial signal.

$$\hat{S}_{A,C=2} = \frac{\frac{x_A}{\sigma_A^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}} \tag{3}$$

To provide a final estimate of the auditory location, the brain can combine the estimates under the two causal structures using various decision functions. In this study, we consider three decision functions for the implicit CI involved in the spatial localization task (for details see Wozny et al. (2010):

According to the "model averaging strategy, the brain combines the two auditory location estimates weighted in proportion to the posterior probability of their underlying causal structure.

$$\begin{aligned}\hat{S}_A = {} & p(C = 1|x_A, x_V)\hat{S}_{AV,C=1} \\ & + (1 - p(C = 1|x_A, x_V))\hat{S}_{A,C=2}\end{aligned} \tag{4}$$

According to the model selection strategy, the brain reports the spatial estimate selectively from the more likely causal structure.

$$\hat{S}_A = \begin{cases} \hat{S}_{AV,C=1} \text{ if } p(C=1|x_A,x_V) > 0.5 \\ \hat{S}_{A,C=2} \text{ if } p(C=1|x_A,x_V) \leq 0.5 \end{cases} \tag{5}$$

According to probability matching, the brain reports the spatial estimate of one causal structure stochastically selected in proportion to its posterior probability.

$$\hat{S}_A = \begin{cases} \hat{S}_{AV,C=1} \text{ if } p(C=1|x_A,x_V) > \alpha, \alpha \sim U(0,1) \\ \hat{S}_{A,C=2} \text{ if } p(C=1|x_A,x_V) \leq \alpha, \alpha \sim U(0,1) \end{cases} \tag{6}$$

Even though probability matching is suboptimal, humans are known to use this strategy in a variety of cognitive tasks (e.g., Gaissmaier & Schooler, 2008). Further, a recent study suggested that human observers use probability matching in audiovisual spatial localization (Wozny et al., 2010).

We also considered two decision strategies for the explicit CI that is involved when generating a binary response (common source vs. independent sources) for the common-source judgment. First, we considered that subjects reported common source when the posterior probability of a common source is greater than the threshold of 0.5 ($\mathrm{CI}_{\mathrm{Th-0.5}}$).

$$\hat{C} = \begin{cases} 1 \text{ if } p(C=1|x_A,x_V) > 0.5 \\ 2 \text{ if } p(C=1|x_A,x_V) \leq 0.5 \end{cases} \tag{7}$$

Second, similar to the probability-matching strategy described above for the spatial localization task, we considered that participants report common source stochastically in proportion to the posterior probability of a common source ($\mathrm{CI}_{\mathrm{Sampling}}$).

$$\hat{C} = \begin{cases} 1 \text{ if } p(C=1|x_A,x_V) > \alpha, \alpha \sim U(0,1) \\ 2 \text{ if } p(C=1|x_A,x_V) \leq \alpha, \alpha \sim U(0,1) \end{cases} \tag{8}$$

Factorially manipulating the decision functions for the spatial localization task and the common-source judgment, we generated a 3 × 2 space of six CI models. We then fitted each of the six CI models jointly to the response data from the spatial localization and the common-source judgment tasks for each subject. Thereby, we obtained model parameters and predictions individually for each subject. The subject-specific parameters and predictions were used to compute the summary indices shown in the figures.

Fitting parameters of the six CI models

The predicted distributions of the auditory spatial estimates (i.e., $p(\hat{S}_A|S_A,\ S_V,\ 1/\sigma_V{}^2)$) and the causal structure estimates (i.e., $p(\hat{C}|S_A,\ S_V,\ 1/\sigma_V{}^2)$) were obtained by marginalizing over the internal variables x_A and x_V. These distributions were generated by simulating x_A and x_V 1,000 times for each of the 100 conditions and inferring $\hat{S}_A$ and $\hat{C}$ from Equations 1–8. To link $p(\hat{S}_A|S_A,\ S_V,\ 1/\sigma_V{}^2)$ to participants' auditory localization responses as discrete button responses, we assumed that participants selected the button that is close to $\hat{S}_A$ and binned the data accordingly. Based on these predicted distributions, we computed the log likelihood of participants' auditory localization and common-source judgment responses. Assuming independence of conditions and task responses, we summed the log likelihoods across conditions and across auditory localization and common-source judgment responses for a particular subject.

To obtain maximum likelihood estimates for the parameters of the models for a particular subject (p_{common}, σ_P, σ_A, $\sigma_V 1$–$\sigma_V 4$ for each of the four levels of visual reliability), we used a nonlinear simplex optimization algorithm as implemented in Matlab's fminsearch function (Matlab R2010b). This optimization algorithm was initialized with 200 different parameter settings that were defined based on a prior grid search. For each subject, we used the parameter setting with the highest log likelihood across the 200 initializations (Supplementary Table S1). This fitting procedure was applied individually to each participant's data set for each of the six CI models to obtain model parameters individually for each subject. We report across-subjects' mean (± standard error) of those parameters as summary statistic indices at the second between-subject or group level.

The model fit was assessed by the coefficient of determination (Nagelkerke, 1991). To identify the optimal model for explaining subjects' data, we compared the CI models using the Bayesian Information Criterion (BIC) as an approximation to the model evidence (Raftery, 1995). The BIC depends on both model complexity and model fit. For comparison at the group level, we summed the individual BICs across subjects (i.e., a fixed-effects approach; see Figure 1C).

In addition, we investigated which decision strategy is most likely given the data separately for implicit CI during spatial localization and for explicit CI during common-source judgments. For this, we partitioned the model space into three (implicit CI) or two (explicit CI) model families according to the 3 × 2 factorial structure of our model space. Thus, we compared the three model families of model selection, model averaging, and probability matching during the spatial localization task. Likewise, we compared the model families of fixed threshold at 0.5 ($\mathrm{CI}_{\mathrm{Th-0.5}}$) and sampling procedure ($\mathrm{CI}_{\mathrm{Sampling}}$) for the common-source judgment. The posterior probability of a model family is simply the sum of the posterior probabilities of each model within this family (Penny et al., 2010; for implementational details see SPM8, www.fil.ion.ucl.ac.uk/spm; Friston et al., 1994).

Comparing human responses to model predictions: Response indices

To inspect whether the most likely CI model can account qualitatively for a participant's response profile, we show participants' responses and the predicted responses of the most likely CI model (Figures 2 through 4). To enable a direct comparison, we processed and formed indices (e.g., the ventriloquist effect) of the model's predicted responses (1,000 trials were simulated per condition) exactly as for the participants' responses (see below). For visualization and didactic purposes, we also present the predicted responses of a traditional forced-fusion model (Alais & Burr, 2004; Ernst & Banks, 2002) that is fitted selectively to the auditory localization data (Figure 2B through D). Yet, the forced-fusion model cannot formally be compared to any of the CI models because it cannot be fitted to the common-source judgment data.

Specifically, we computed and presented the following response indices: For the common-source judgment, we show the percentage of common-source judgments (Figure 2A). For the spatial localization task, we present the absolute visual bias on the perceived auditory location, which is computed as the deviation of the responded auditory location from the true auditory location (i.e., $A_{Resp} - A_{Loc}$; Figure 2B). Moreover, we show the ventriloquist effect (i.e., the relative visual bias on the perceived auditory location) computed as $VE = (A_{Resp} - A_{Loc}) / (V_{Loc} - A_{Loc})$ with A_{Resp} = mean auditory localization response for a given condition, A_{Loc} = auditory signal location and V_{Loc} = visual signal location (Figures 2C, 3A, B). However, both the absolute and relative visual biases will be greater than zero, even when the visual signal has no influence on the auditory signal and vice versa. This is because participants predominantly make erroneous localization responses towards more central positions in particular for extreme positions where they do not have the choice to respond to more eccentric positions. To account for these spatial response biases, we adjusted A_{Loc} and V_{Loc} with a linear regression approach across all congruent trials irrespective of the level of visual reliability in a subject-specific fashion. In other words, we replaced the true A_{Loc} and V_{Loc} in the crossmodal bias equations with the A_{Loc} and V_{Loc} predicted based on participants' responses during the congruent conditions. Based on simulation results, this adjustment procedure ensures that the crossmodal bias approximately measures the true underlying bias. Hence, the adjusted crossmodal bias reliably reflects the influence of a visual signal on auditory localization responses.

Finally, we evaluated the localization variability of the auditory localization responses as quantified by their variance (Figures 2D, 3C, D). Note, however, that interpreting the variance for binned responses requires caution in particular when the bins are not equally spaced (e.g., eccentric vs. central bins), and unequal binning may affect visual reliability levels differently.

Each of these response indices was computed for each of the 100 conditions in our 5 (auditory locations) × 5 (visual locations) × 4 (visual reliability levels) factorial design. We then reorganized these 100 conditions according to audiovisual spatial disparity and visual reliability (Figure 2A through D). For this, we averaged the indices across all combinations of audiovisual locations at a particular level of spatial disparity and visual reliability (note, this averaging procedure is valid under the assumption that the visual bias is similar across different positions along the azimuth).

In addition, we analyzed the ventriloquist effect and localization variability as a function of common-source judgment by categorizing subjects' spatial localization responses according to whether participants responded common or separate source on those trials. If we treated the subjects' common-source judgment as an independent factor, the factor induced an unbalanced distribution of trials across conditions, such that only few subjects had trials for all combinations of the factors' spatial disparity, visual reliability, and common-source judgment. Thus, for computing the ventriloquist effect, this analysis would have been limited to 13 subjects. Moreover, for computing the localization variability, the analysis would have been limited even to only one single subject, as the computation of localization variability requires at least two trials per condition. When separating for common- versus independent-source judgments, we therefore analyzed and presented the indices pooled either over the factor audiovisual disparity (Figure 3A, C) or visual reliability (Figure 3B, D). To ensure that the effects in the reliability × common-source design could be evaluated unconfounded by differences in disparity, we included only disparity levels that were present in all conditions for the remaining 4 (reliability) × 2 (common source) design in a particular subject. Likewise, when evaluating the effects in the disparity × common-source design, we included only those reliability levels that were present in all conditions for this design. This procedure enabled us to include full data sets from at least 25 subjects for the ventriloquist effect and the localization variability in both designs.

Model-free analysis of the common-source judgments, ventriloquist effect, and localization variability

The common-source judgments were characterized in terms of the percentage perceived common source as a function of reliability and audiovisual spatial disparity. We then fitted Gaussian functions (i.e., a height, width, and mean parameter) to the percentage perceived common source as a function of the signed audiovisual disparity separately for each level of reliability (Figure 2A). The effects of visual reliability on the height and width parameters were each assessed in a one-way repeated measures ANOVA.

The spatial localization responses were analyzed in terms of the relative audiovisual bias (i.e., ventriloquist effect) and the localization variability (i.e., variance). Both the ventriloquist effect and the localization variability were analyzed in separate visual reliability (four levels) × spatial disparity repeated measures ANOVA. The factor spatial disparity had five levels for the localization variability, but only four levels for the ventriloquist effect as the computation of the ventriloquist effect requires a disparity greater zero.

We report Greenhouse-Geisser corrected p values and degrees of freedom. Effect sizes were reported as η^2.

Results

Comparison of the CI models

All six CI models were fitted jointly to participants' auditory localization and common-source judgment responses and explained >64% of the variance ($R^2 >$ 64%; cf. Supplementary Table S1). The smaller coefficient of determination results from the fact that the CI models in the current study included only seven parameters to explain the variance across 100 conditions (compared to four parameters explaining 35 conditions in Koerding et al., 2007, and Wozny et al., 2010).

Next, we identified the CI model that maximally accounted for participants' responses jointly during the auditory localization and the common-source judgment tasks by comparing the relative model evidence (BIC) of the CI models in our 3 × 2 model space (Figure 1C). In the winning model, participants used the following decision strategies: For implicit CI in the auditory localization task, participants used model averaging as a decision strategy. Hence, they combined the spatial estimates under the two causal structures weighted by the posterior probabilities of each causal structure. For explicit CI during common-source judgments, participants reported "common source" if the posterior probability was larger than an optimal threshold of 0.5 ($\text{CI}_{\text{Th}-0.5}$). The BIC difference between this model and the second best model was 61.7, which is generally considered as very strong evidence for the winning model (Raftery, 1995). Likewise, family inference (Penny et al., 2010) demonstrated the highest posterior probability for the model averaging strategy for the auditory localization task and the threshold ($\text{CI}_{\text{Th}-0.5}$) decision strategy for the common-source judgment task (cf. Figure 1C). In short, for both implicit and explicit CI, we did not observe evidence for a sampling strategy as was previously reported (Wozny et al., 2010).

The parameters for the visual standard deviation of the winning CI model approximately matched the standard deviation of the Gaussian cloud across the four visual reliability levels (Supplementary Table S1). Thus, visual standard deviation parameters estimated in bisensory conditions appear larger as compared to parameters estimated in unisensory conditions that are found to be smaller than the nominal standard deviation of Gaussian stimuli (see Supplementary Table S4 and Alais & Burr, 2004). The auditory standard deviation parameter was comparable to the lowest visual standard deviation. The common-source prior was approximately 0.5 indicating that participants a priori assumed that signals were equally likely to come from common or independent sources.

In addition to the CI model we have also fitted the previously reported "coupling prior model" to participants' auditory localization responses (note, this model cannot be fitted to participants' common-source judgments). The coupling prior model accounts for partial integration by imposing a Gaussian coupling prior that controls the extent to which sensory signals are integrated (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006; see Supplementary Methods). As previously reported (Koerding et al., 2007), the CI model outperformed the coupling prior model (Supplementary Table S2; note that we fitted the CI model only to the participants' auditory localization responses for this comparison). In particular, the coupling prior model could not account for the fact that multisensory integration breaks down progressively with increasing spatial discrepancy (i.e., the nonlinearity observed for the absolute visual bias as a function of spatial discrepancy; see Figure 2B; cf. Supplementary Discussion).

To further investigate whether the winning CI model qualitatively replicated participants' response profile, we compared participants' common-source judgments and auditory localization responses with the response predictions obtained from the CI model. More specifically, we show the common-source judgments, the absolute and relative visual bias (i.e., ventriloquist effect), and the variability (i.e., variance) during the

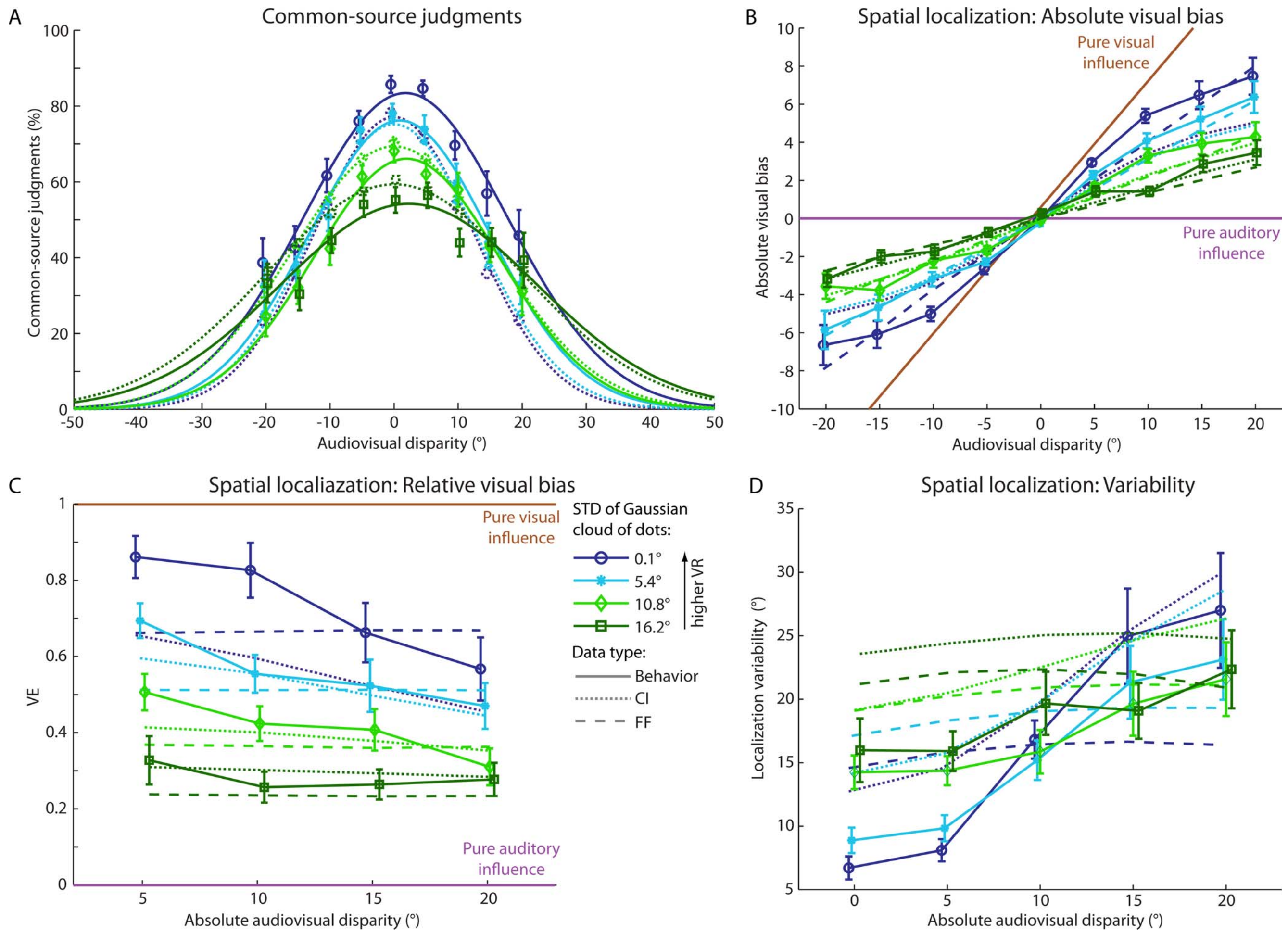

Figure 2. Behavioral responses and the models' predictions (pooled over common-source judgments). The figure panels show the behavioral data (across subjects' mean ± *SEM*; solid lines) and the predictions of the winning CI model (dotted lines) and the forced-fusion model (FF; dashed lines) as a function of visual reliability (VR; color coded) and audiovisual disparity (shown along the *x*-axis). (A) Percentage of common-source judgments. (B) Absolute spatial visual bias, $A_{Resp} - A_{Loc}$. (C) Relative spatial visual bias (i.e., the ventriloquist effect $VE = (A_{Resp} - A_{Loc}) / (V_{Loc} - A_{Loc})$). In panels (B) and (C), the absolute and relative spatial visual biases are also shown for the case of pure visual or pure auditory influence for reference. (D) Localization variability (i.e., variance) of the behavioral and models' predicted responses.

auditory localization task computed from participants' responses and the model's predicted responses.

Analysis of common-source judgments

The common-source judgments peaked at zero and decayed as a function of audiovisual disparity according to a Gaussian bell-shaped function ($R^2 > 86\%$, explained variance averaged across the levels of visual reliability; Figure 2A). A higher visual reliability significantly increased the height (effect of visual reliability on height parameter, $F(2.5, 61.9) = 32.995$, $p < 0.001$, $\eta^2 = 0.569$) and marginally changed the width of the Gaussian (effect of visual reliability on width parameter, $F(2.2, 54.2) = 2.606$, $p = 0.079$, $\eta^2 = 0.094$). The width of the Gaussian can be interpreted as an index for the width of the audiovisual integration window when it is judged explicitly in common-source judgments by participants. Our results demonstrate that participants were generally more likely to infer a common source at high relative to low visual reliability. Moreover, the slopes of the Gaussian functions were greater at high visual reliability, indicating that high visual reliability rendered spatial disparity a more informative cue for discriminating between common source and independent sources.

Critically, the CI model qualitatively replicated, though slightly underestimated, the effect of visual reliability (cf. Figure 2A). Thus, the model predicted fewer common-source judgments for high visual reliability and more frequent common-source judgments for low visual reliability.

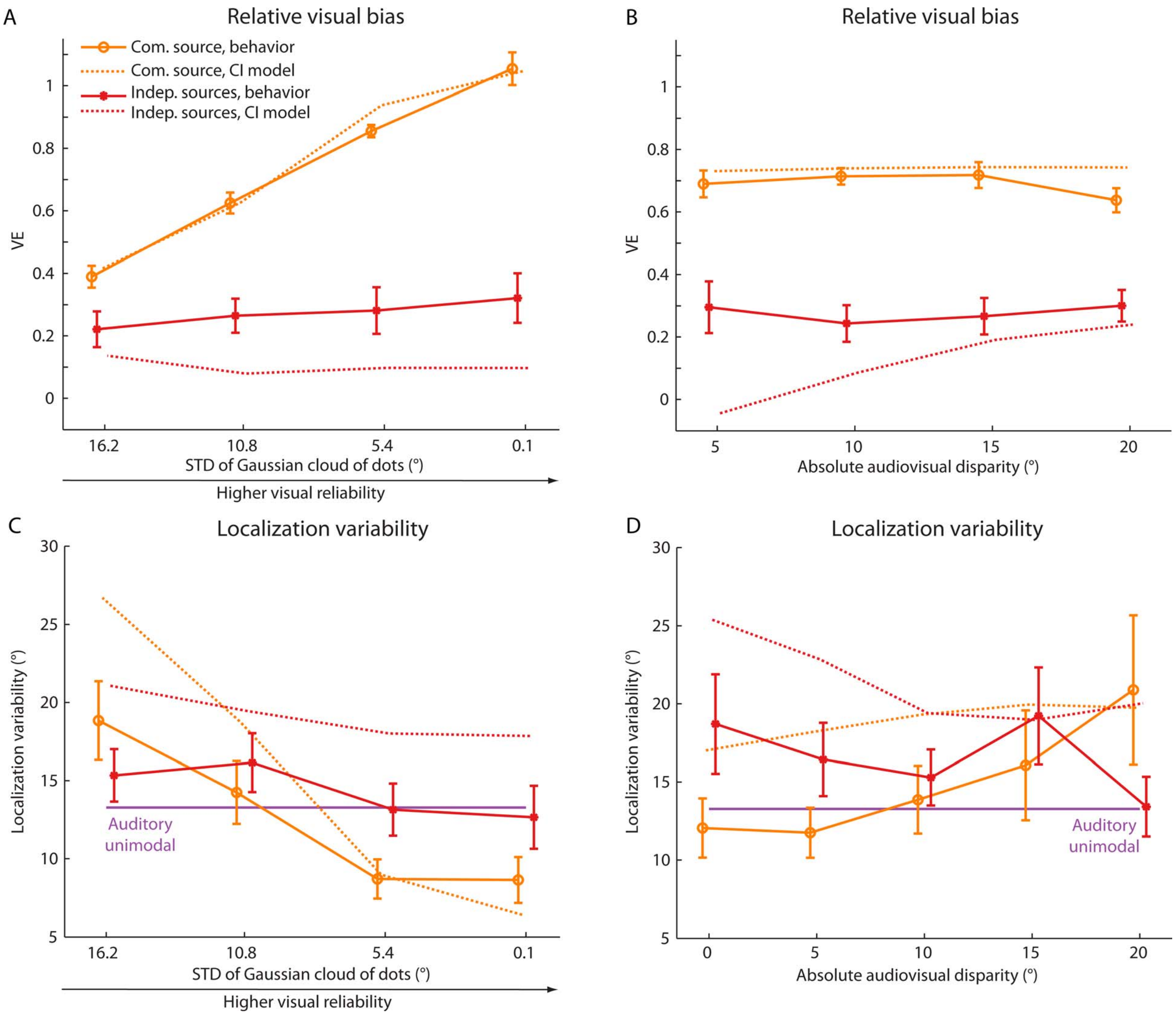

Figure 3. Behavioral responses and the model's predictions (separated according to common-source decisions). The figure panels show the behavioral data (across subjects' mean ± *SEM*; solid lines) and the predictions of the winning CI model (dotted lines) as a function of visual reliability (left, A, C) and audiovisual disparity (right, B, D). The ventriloquist effect (VE; A, B) and localization variability (i.e., variance; C, D) are shown separately for trials where common or independent sources were inferred. Localization variability for unimodal auditory trials is shown as a solid line for reference.

Analysis of the visual bias and localization variability, irrespective of common-source judgments

Visual bias

The visual influence on perceived auditory location was evaluated using the absolute visual bias (i.e., $A_{Resp} - A_{Loc}$; Figure 2B) and the relative visual bias also referred to as ventriloquist effect (i.e., VE = $(A_{Resp} - A_{Loc}) / (V_{Loc} - A_{Loc})$; Figure 2C). Both indices are qualitatively in line with the predictions of Bayesian CI. Thus, the absolute visual bias increased nonlinearly. This indicated that audiovisual integration breaks down when large spatial discrepancies render a common source unlikely. Likewise, for the relative visual bias (i.e., ventriloquist effect; Figure 2C), we observed not only a main effect of visual reliability, $F(1.6, 38.7) = 46.147$, $p < 0.001$, $\eta^2 = 0.649$, as predicted by forced-fusion models (Alais & Burr, 2004; Ernst & Banks, 2002), but also a main effect of absolute disparity, $F(1.5, 37.8) = 21.339$, $p < 0.001$, $\eta^2 = 0.460$. Again as predicted by Bayesian CI, the ventriloquist effect is reduced for large spatial discrepancies when it is unlikely that the two signals come from a common source.

Critically, we also observed a significant interaction between visual reliability and spatial disparity (Figure 2C; interaction effect of visual reliability and absolute disparity, $F(5.5, 138.1) = 3.511$, $p = 0.004$, $\eta^2 = 0.123$). This interaction emerged because visual reliability changes the height and nonsignificantly the width of the audiovisual integration window. In other words, the

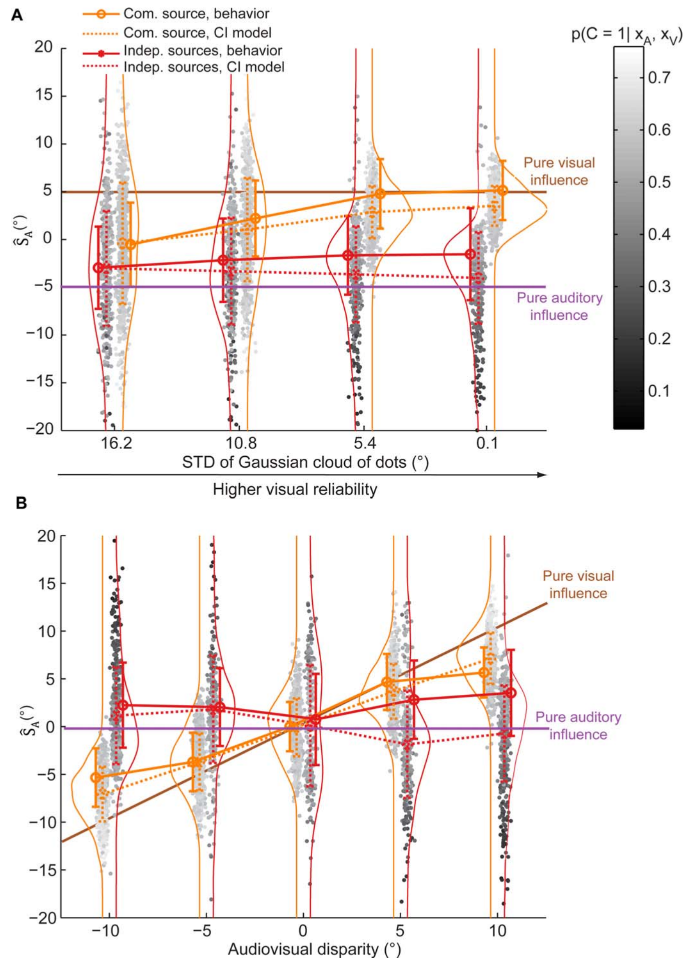

Figure 4. Distributions of perceived auditory locations ($\hat{S}A$, along the *y*-axis) simulated by the winning Bayesian CI model as a function of visual reliability (top, A) and audiovisual disparity (bottom, B). The gray tone of each dot encodes the posterior common-source probability ($p[C = 1] \mid xA, xV$). For each level of reliability or disparity, the dots are assigned to one of two clouds depending on whether the posterior probability of a common source is smaller than 0.5 (i.e., left cloud) or larger than 0.5 (i.e., right cloud). The probability densities of $\hat{S}A$ are summarized as line plots separately for the two clouds of perceived auditory locations. Further, the mean and the standard deviation of the predicted responses (dotted lines in red and yellow) and the observed behavioral responses (solid lines in red and yellow) are shown. In (A), the visual and auditory signal locations are fixed at 5° and −5°, respectively (i.e., a constant spatial disparity of 10°). In (B), the visual reliability is fixed at 5.4° and the auditory signal location is fixed at 0°. For reference, the auditory responses in case of pure visual or pure auditory influence are shown as solid lines.

shape of the Gaussian functions characterizing the common-source judgments (cf. Figure 2A) indicates that less spatial disparity is needed for the brain to infer that audiovisual signals should be segregated at high visual reliabilities. These sharper audiovisual integration windows also make the ventriloquist effect decrease faster with spatial disparity when the visual signals are reliable.

Localization variability

The central benefit of multisensory integration is that it produces audiovisual estimates that are more reliable (i.e., less variable; Ernst & Banks, 2002). Indeed, for small spatial disparities we observed that the auditory localization variability decreased with higher visual reliability (Figure 2D). However, as predicted by Bayesian CI, this reduction in localization variability was no longer observed for large spatial discrepancies indicating a breakdown of audiovisual integration (i.e., interaction effect of visual reliability and absolute disparity, $F(4.6, 115.0) = 3.229$, $p = 0.011$, $\eta^2 = 0.114$; and a main effect of absolute disparity, $F(1.8, 45.6) = 20.491$, $p < 0.001$, $\eta^2 = 0.450$). Note, however, that the CI model generally overestimated the localization variability by a constant amount as has previously been reported (Körding & Tenenbaum, 2006; Natarajan, Murray, Shams, & Zemel, 2009).

Analysis of visual bias and localization variability, dependent on common-source judgments

Visual bias

Next, we investigated participants' auditory localization responses and the predictions of the CI model separately for trials on which participants inferred common or independent sources (Figure 3). To illustrate how some of these effects on bias and localization variability emerge from splitting the localization response distributions according to the posterior common-source probability, we have also added Figure 4 that shows the predicted distributions of the localization responses (along the *y*-axis) and posterior common-source probability (gray-tone coded) as a function of visual reliability (Figure 4A) and spatial disparity (Figure 4B).

As expected under Bayesian CI, we observed an overall larger ventriloquist effect that progressively increased with higher visual reliability when a common source was inferred (Figure 3A). By contrast, when no common source was inferred, the ventriloquist effect was only negligibly influenced by visual reliability. Likewise, once the outcome of explicit CI was taken into account, the effect of spatial disparity (cf. Figure 2C) was nearly abolished, and the ventriloquist effect differed approximately by a constant when common and independent sources were inferred (Figure 3B). While this response profile is approximately in line with Bayesian CI, the CI model (under Gaussian assumptions) would have predicted a repulsion effect for trials when independent sources were inferred (cf. Supplementary Figure S2B). In other words, on those trials the perceived and reported auditory location should have been shifted away from the visual signal location. While a repulsion effect has indeed previously been shown for human localization responses (Koerding et al., 2007; Wallace et al., 2004), our study did not replicate this effect (Supplementary Figure S2A).

The reasons for the differences in response profiles between the current study and Wallace et al. (2004) are not clear. Potentially, a repulsion effect in our experiment was not observed because the visual stimulus was a cloud of dots rather than an LED flash or the sounds were delivered via headphones. Further, Wallace et al. (2004) did not vary visual reliability across trials. This is important because the CI model predicts the strongest repulsion effect when auditory and visual signals are highly reliable. Second and most prominently, our study used discrete responses via a key press, while Wallace et al. (2004) enabled a continuous response using a laser pointer. Even though the CI model predicts a repulsion effect irrespective of discrete or continuous response options, imposing discrete response options may have altered participants' decisional strategies in ways that are not yet fully accommodated by the CI model. For instance, discrete response options may have introduced response biases such as equalization of responses or transformed the spatial estimation task into a categorization task (e.g., learning the mapping from audiovisual spatial signals onto five response categories). In line with this conjecture, we observed a pronounced repulsion effect in a separate sample of four subjects that employed a joystick (without visual feedback) to indicate the perceived sound location (see Supplementary Methods and Results; Supplementary Figure S4). Yet, as the number of subjects was rather small, it seems premature to draw final conclusions. Furthermore, the response distributions when using a joystick were non-Gaussian, thereby violating the Gaussian assumptions of the CI model as implemented in the main paper.

Interestingly, a recent study comparing the predictions made by the CI model when implemented with Gaussian or heavy-tailed likelihood distributions demonstrated that Gaussian assumptions are critical for the emergence of the repulsion effect (Natarajan et al., 2009). Even though the CI model with heavy-tailed distributions provided better predictions for localization variability, surprisingly it did not predict the repulsion effect.

In summary, the conditions under which the repulsion effect emerges in experimental settings and modeling approaches require further investigation. To our knowledge, the repulsion effect has so far been reported only in one single experimental study by Wallace et al. (2004) that has been referred to and modeled by Koerding et al. (2007). Further, we have now observed a repulsion effect anecdotally in the four subjects that indicated their location judgment via a joystick (see Supplementary Results; Supplementary Figure S4). Conversely, from the modeling perspective further work is needed to determine the critical assumptions (e.g., Gaussian vs. heavy-tailed likelihood distributions) under which the repulsion effect emerges in the CI model.

Localization variability

Not surprisingly, the profile of auditory localization variability also depended on the outcome of participants' common-source judgment. As expected under the CI model, auditory localization variability was only negligibly influenced by visual reliability when participants inferred independent sources and segregated information. By contrast, the auditory localization variability was smaller than during unisensory conditions at least for high visual reliability when participants inferred a common source (Figure 3C) and benefitted from audiovisual integration.

Likewise, the effect of spatial disparity on localization variability depended on the outcome of participants' common-source judgments (Figure 3D). Interestingly, for both participants' and model's responses, the localization variability was decreased for small spatial disparities when a common source was inferred. Yet, it increased for spatial disparities when independent sources were inferred. This effect can be explained by the fact that participants infer independent sources predominantly when the observed visual signal is located far away from the auditory signal, either to its left or right. Thus, when no common source is inferred for spatially congruent signals, the auditory localization responses come from a bimodal distribution leading to an increase in localization variability (see Figure 4B).

In conclusion, the behavioral response profile observed in the current study suggests that participants' explicit common-source judgment partially separates the spatial localization responses into two classes: When a common source is inferred, auditory localization responses conform approximately to predictions of the forced-fusion model. In other words, participants weight the sensory signals according to their reliability (Figure 4A) in a linear fashion (Figure 4B). By contrast, when no common source is inferred, participants responded predominantly based on the auditory signal approximately as predicted by a segregation model where signals are processed independently. Yet, while the explicit common-source judgment in our study provided only the binary response options "common versus separate" sources, the model averaging strategy weights the spatial estimates of the two causal structures by their continuous posterior probability. To relate explicit and implicit common-source judgments even more closely, a future study may therefore provide participants with a continuous response option (e.g., a rating scale) for the common-source judgment (Lewald & Guski, 2003).

Discussion

The current study investigated the decision strategies that observers use for inferring the causal structure of audiovisual spatial signals when probed implicitly in an auditory localization or explicitly in a common-source judgment task. Given the critical role of sensory reliability in integration within (Jacobs, 1999; Knill & Saunders, 2003; Oruc, Maloney, & Landy, 2003) and across the senses (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; Yuille & Buelthoff, 1996), we evaluated the CI model on psychophysics data that included multiple levels of visual reliability.

It is well established that sensory signals should only be integrated when they are close in time, space, and structure (Lee & Noppeney, 2014; Lewald & Guski, 2003; Lewis & Noppeney, 2010; Noppeney, Ostwald, & Werner, 2010; Roach, Heron, & McGraw, 2006; Rohe & Noppeney, 2015; Slutsky & Recanzone, 2001; Wallace et al., 2004; Welch & Warren, 1980). Recently, this problem has been framed within probabilistic Bayesian CI (Knill, 2007; Koerding et al., 2007; Sato, Toyoizumi, & Aihara, 2007; Shams & Beierholm, 2010), where a response during implicit (i.e., in spatial localization) and explicit (i.e., common-source judgments) CI tasks can be formed based on several decision strategies (Wozny et al., 2010).

Our results show that human observers employ model averaging as a decision strategy for implicit CI in auditory localization. In other words, they obtain an auditory localization estimate by combining the spatial estimates under the two causal structures weighted by their posterior probabilities. The model averaging strategy minimizes the squared error of signal localizations and simultaneously accounts for the uncertainty of the underlying causal structure. By contrast, in a previous study the majority of participants used non-optimal probability matching for auditory localization (Wozny et al., 2010). These inconsistencies may arise from differences in the visual spatial signals (i.e., Gaussian cloud vs. LED) or the number of visual reliability levels (i.e., four vs. one) across the two experiments. Further, complex dual task effects (Stanovich & West, 2000; Stocker & Simoncelli, 2007) may explain the differences as the current design combined auditory localization and common-source judgment, while the previous study included auditory and visual localization tasks. Thus, Stocker and Simoncelli (2007) have previously shown that prior categorization of motion direction may affect subsequent estimation of motion direction by biasing the perceived motion direction away from the decision boundary applied in the categorization task (though see Jazayeri & Movshon, 2007, for an alternative explanation). They have explained this repulsion effect in a model where the initial categorization task makes participants condition their estimation of the motion direction under the assumption that the motion direction comes indeed from the selected motion direction category. In other words, the categorization task influenced the estimation task because of prior model selection. While in our study model averaging fitted participants' responses better than model selection (despite the categorical common-source judgment task), we cannot exclude that the dual task context may have introduced other types of strategic adjustments and biases.

For explicit CI probed in the common-source judgment task, we observed that participants reported a common source if the common-source posterior probability was larger than 0.5. Thus, neither for implicit CI during spatial localization nor for explicit CI during common-source judgments did participants in our study employ suboptimal sampling strategies where they selected each causal structure stochastically in proportion to its posterior probability.

Moving beyond previous modeling efforts (Beierholm et al., 2009; Koerding et al., 2007; Sato et al., 2007; Wozny et al., 2010), we validated Bayesian CI models on two psychophysics data sets that included two (see Supplementary Experiment 1) or four levels of visual reliabilities. This is critical, because according to the Bayesian CI model, sensory reliability influences multisensory integration via two distinct mechanisms: CI and reliability-weighted integration. As expected under Bayesian CI, visual reliability sharpened the audiovisual integration window (Figure 2A; Supplementary Figure S3A). Participants were better at discriminating whether sensory signals came from a common or two independent sources when the visual signals were highly reliable. Further, lower visual noise (i.e., the inverse of visual reliability) generally reduced the probability that observers perceived the two signals as coming from common sources. In other words, visual noise biased participants' common-source judgments (for related sensory noise-dependent biases in visual categorization tasks, see Qamar et al., 2013). As predicted by both forced fusion (Alais & Burr, 2004; Ernst & Banks, 2002) and CI models (Koerding et al., 2007), high visual reliability also increased the influence of the visual signal on the perceived auditory location leading to a larger ventriloquist effect (Figure 2C; Supplementary Figure S3C). Yet, in contradiction to the forced-fusion model, spatial ventriloquism broke down for greater spatial disparity when it is unlikely that audiovisual signals come from a common source. Moreover, we observed a significant interaction between reliability and spatial disparity. In other words, high visual reliability amplified the decay in ventriloquism with greater spatial disparity by sharpening the integration window. Likewise, the localization variability depended on both spatial disparity and visual reliability in an interactive fashion (Figure 2D; Supplementary Figure S3D).

In summary, visual reliability influenced the ventriloquist effect and localization variability via two interacting mechanisms: (a) sharpening of the integration window via CI and (b) reliability-weighted integration in the case of a common source.

These two hierarchically organized mechanisms can be partially dissociated by separating localization responses depending on whether or not participants perceived a common source. Indeed, accounting for CI by separating trials according to participants' common-source judgments largely abolished the effect of spatial disparity on the ventriloquist effect and localization variability both for human responses and the predictions of the CI model (cf. Figure 2C, D vs. Figures 3B, D, 4B). Likewise, the effect of reliability on spatial ventriloquism and localization variability emerged predominantly when a common source was inferred (Figure 3A, C and Figure 4A). Collectively, these results suggest that reliability-weighted integration as a special case of multisensory integration is predicated on CI. Yet, when separating localization responses according to the outcome of the common-source judgments, we still observed small effects of spatial discrepancy on spatial bias and localization variability. In particular, the localization variability increased for small spatial discrepancies when independent sources were inferred. As shown in Figure 4, this surprising effect emerges because independent sources are inferred if the auditory percept is distant from the visual signal, either to the left or to the right, leading to a bimodal response distribution (Figure 4B).

In summary, the Bayesian CI model explains how spatial disparity and sensory reliability shape common-source judgments, the ventriloquist effect, and localization variability. First, it models how signal integration at small spatial disparities turns into signal segregation at large spatial disparities during spatial localization. As previously shown by Shams and Beierholm (2010), there is a range of models that can account for this nonlinear relationship between localization bias and spatial disparity or cue conflict by assuming mixture distributions as a prior (e.g., two Gaussians with different variances or a Gaussian plus a uniform distribution; Roach et al., 2006; Sato et al., 2007) or heavy-tailed likelihood distributions (Girshick & Banks, 2009; Natarajan et al., 2009). Second, the Bayesian CI model also allows us to predict participants' common-source judgments by explicitly modeling the underlying two potential causal structures. Thus, by explicitly reconstructing the causal structure of the environment, the Bayesian CI model can simultaneously model categorical responses for common-source judgments and continuous estimates for the spatial localization task. Yet, while the Bayesian CI model qualitatively predicts participants' responses on both tasks, there are several quantitative mismatches that require further investigation.

First, the Bayesian CI model overestimated the localization variability in particular for small spatial disparities in the two experiments with four and two visual reliability levels (see Figure 2D and Supplementary Figure S3D). However, in Supplementary Experiment 1 (see Supplementary Material) this mismatch occurred only when the visual and auditory variances were estimated together with the other parameters during multisensory conditions. By contrast, when the sensory variance parameters were estimated in unisensory conditions and then entered as known when fitting the remaining parameters (i.e., the common-source prior) with the multisensory conditions, the localization variability was relatively well predicted for the small spatial disparity trials. This profile suggests that the sensory variance may depend on the attentional context. In classical forced-fusion paradigms (Alais & Burr, 2004; Ernst & Banks, 2002), the auditory and visual signals were only shown with a small unnoticeable conflict, so that participants should ideally always attend to and integrate the two sensory signals. By contrast, in our ventriloquist paradigm participants were presented with audiovisual signals of variable spatial disparities. From a more cognitive perspective, the current paradigm can be considered an intersensory selective-attention paradigm where participants should arbitrate between sensory integration and segregation depending on spatial disparity. In such an intersensory selective-attention paradigm, the sensory variance may be influenced by attentional modulation as a function of spatial disparity. For instance, during auditory localization the unattended variance of the internal visual representation may depend on the spatial discrepancy of the task-relevant auditory signal. Future experiments are needed to better understand how the model needs to be extended to account for these more complex and dynamic interdependencies between attentional context and spatial discrepancy.

Second, the Bayesian CI model slightly underestimated the percentage of signals judged as coming from a common source for high visual reliability and overestimated those for low visual reliability trials. While this mismatch may potentially be related to the problem of sensory variance estimation discussed above, future studies are needed to better understand this mismatch.

Finally, as already discussed in detail in the results section, we observed a repulsion effect for signals judged as coming from independent sources only in the small sample of four subjects that indicated their auditory localization responses via a joystick (see Supplementary Experiment 2; Supplementary Figure S4). To our knowledge, the repulsion effect has so far been described only in a single study by Wallace et al. (2004) where participants indicated their spatial responses via a laser pointer. Further, a recent modeling study suggested that the repulsion effect emerges only when Gaussian but not when heavy-tailed distributions are assumed for the sensory likelihood distribution (Natarajan et al., 2009). Thus, future studies need to investigate the experimental and modeling factors that determine whether a repulsion effect emerges in observed responses (e.g., discrete button choices vs. continuous laser pointer) and model predictions.

Research into the neural basis of multisensory integration has focused predominantly on the special case of reliability-weighted integration under forced-fusion assumptions (Beauchamp, Pasalar, & Ro, 2010; Fetsch, DeAngelis, & Angelaki, 2013; Helbig et al., 2012). For instance, very elegant neurophysiology work in macaque has demonstrated that single neurons integrate sensory inputs linearly weighted by their reliability (Morgan, Deangelis, & Angelaki, 2008) in line with theories of probabilistic population coding (Ma, Beck, Latham, & Pouget, 2006). Furthermore, in a visuo-vestibular heading task decoding of neuronal activity in the dorsal medial superior temporal area (MSTd) mostly accounted for the sensory weights that the nonhuman primates employed at the behavioral level (Fetsch, Pouget, DeAngelis, & Angelaki, 2012). A recent fMRI study suggests that a cortical hierarchy performs Bayesian CI for spatial localization by representing multiple spatial estimates, which are closely related to forced-fusion (e.g., $\hat{S}_{AV,C=1}$), full-segregation (e.g., $\hat{S}_{AV,C=2}$), and Bayesian CI (e.g., $\hat{S}_A$; Rohe & Noppeney, 2015). Future neurophysiology studies in primates are needed to investigate how single neurons or populations of neurons implement Bayesian CI.

Conclusions

The current study demonstrates that Bayesian CI is fundamental for multisensory integration in our uncertain natural environment. Sensory reliability critically shapes multisensory integration via two distinct mechanisms. First, it determines CI by sharpening the integration window. Second, it determines the relative weights of the sensory inputs in the integration process under the assumption of a common source.

Keywords: cue combination, perceptual inference, multisensory integration, Bayesian Causal Inference, sensory reliability

Acknowledgments

This work was funded by the Max Planck Society and the European Research Council (ERC-multsens grant number 309349). We thank Ulrik Beierholm for helpful comments on a previous version of the manuscript.

Commercial relationships: none.
Corresponding author: Tim Rohe.
Email: tim.rohe@tuebingen.mpg.de.
Address: Max Planck Institute for Biological Cybernetics, Tuebingen, Germany.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). *The CIPIC HRTF database*. Paper presented at the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, New Paltz, New York.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America, A: Optics, Image Science, & Vision*, *20*(7), 1391–1397.

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in Systems Neuroscience*, *4*, 25.

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5):23, 1–29, http://www.journalofvision.org/content/9/5/23, doi:10.1167/9.5.23. [PubMed] [Article]

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5):2, 554–564, http://www.journalofvision.org/content/6/5/2, doi:10.1167/6.5.2. [PubMed] [Article]

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105–131). New York: Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–154.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2*(4), 189–210.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*(3), 416–422.

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11):7, 1013–1023, http://www.journalofvision.org/content/5/11/7, doi:10.1167/5.11.7. [PubMed] [Article]

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *Journal of Vision*, *9*(9):8, 1–20, http://www.journalofvision.org/content/9/9/8, doi:10.1167/9.9.8. [PubMed] [Article]

Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., … Noppeney, U. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage*, *60*(2), 1063–1072.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629.

Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, *446*(7138), 912–915.

Kleiner, M., Brainard, D., & Pelli, D. (2007). "What's new in Psychtoolbox-3?" *Perception 36 ECVP Abstract Supplement*. Presented at ECVP 2007, Arezzo, Italy.

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7):5, 1–24, http://www.journalofvision.org/content/7/7/5, doi:10.1167/7.7.5. [PubMed] [Article]

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558.

Koerding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), e943.

Körding, K. P., & Tenenbaum, J. B. (2006). Causal inference in sensorimotor integration. *Advances in Neural Information Processing Systems*, *2006*, 737–744.

Lee, H., & Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Current Biology*, *24*(8), R309–R310.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Research: Cognitive Brain Research*, *16*(3), 468–478.

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience*, *30*(37), 12329–12339.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Morgan, M. L., Deangelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–673.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.

Natarajan, R., Murray, I., Shams, L., & Zemel, R. (2009). Characterizing response behavior in multisensory perception with conflicting cues. *Advances in Neural Information Processing Systems, 2009*.

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *Journal of Neuroscience*, *30*(21), 7434–7446.

Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing families of dynamic causal models. *PLoS Computational Biology*, *6*(3), e1000709.

Qamar, A. T., Cotton, R. J., George, R. G., Beck, J. M., Prezhdo, E., Laudano, A., . . . Ma, W. J. (2013). Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proceedings of the National Academy of Sciences*, *110*(50), 20332–20337.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings: Biological Science*, *273*(1598), 2159–2168.

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation*, *19*(12), 3335–3355.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Science*, *14*(9), 425–432.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, *23*(5), 645–665, 665–726.

Stocker, A. A., & Simoncelli, E. P. (2007). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems*, *2007*, 1409–1416.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, *94*(1), 111–123.

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, *85*(2), 868–878.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, *6*(8), e1000871.

Yuille, A. L., & Buelthoff, H. H. (1996). *Bayesian decision theory and psychophysics*. New York: Cambridge University Press.