

Bayesian priors are encoded independently from likelihoods in human multisensory perception

Ulrik R. Beierholm

Gatsby Computational Neuroscience Unit, UCL, London, UK



Steven R. Quartz

Division of Humanities and Social Sciences, Caltech,
Pasadena, CA, USA



Ladan Shams

Department of Psychology, UCLA, Los Angeles, CA, USA



It has been shown that human combination of crossmodal information is highly consistent with an optimal Bayesian model performing causal inference. These findings have shed light on the computational principles governing crossmodal integration/segregation. Intuitively, in a Bayesian framework priors represent *a priori* information about the environment, i.e., information available prior to encountering the given stimuli, and are thus not dependent on the current stimuli. While this interpretation is considered as a defining characteristic of Bayesian computation by many, the Bayes rule *per se* does not require that priors remain constant despite significant changes in the stimulus, and therefore, the demonstration of Bayes-optimality of a task does not imply the invariance of priors to varying likelihoods. This issue has not been addressed before, but here we empirically investigated the independence of the priors from the likelihoods by strongly manipulating the presumed likelihoods (by using two drastically different sets of stimuli) and examining whether the estimated priors change or remain the same. The results suggest that the estimated prior probabilities are indeed independent of the immediate input and hence, likelihood.

Keywords: Bayesian inference, causal inference, priors, likelihoods

Citation: 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, 1–9, <http://journalofvision.org/9/5/23/>, doi:10.1167/9.5.23.

Introduction

Bayesian inference is a statistically optimal way of combining information sources about a hidden property given noisy/uncertain environment and sensory representations. A number of studies have examined whether human multisensory information integration follows the rules specified by Bayes law. Generally, the evidence indicates that it does (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Beierholm, Kording, Shams, & Ma, 2008; Ernst & Banks, 2002; Ghahramani, 1995; Körding et al., 2007; Shams, Ma, & Beierholm, 2005; van Beers, Sittig, & Gon, 1999). Similar studies have been performed on cues within modalities (e.g. texture and motion, Jacobs, 1999) or stereo and shading (Bülthoff & Mallot, 1988), all demonstrating a close consistency between human perception and Bayesian inference.

Furthermore, recent studies have shown that human multisensory perception, not confined to integration but spanning the entire range from integration to segregation, is remarkably consistent with a Bayesian observer performing causal inference (Körding et al., 2007, see also Bresciani, Dammeier, & Ernst, 2006; Roach, Heron, & McGraw, 2006; Shams et al., 2005). Together, these results suggest that human multisensory perception is Bayes-optimal.

In a Bayesian framework, perceptual decisions are based upon the posterior probability distribution, which is obtained by combining the likelihood distributions (of for example, visual stimulus ν) and prior distributions: $\text{Posterior}(\nu) \propto \text{Likelihood}(\nu) * \text{Priors}$. Demonstrating that a task is performed in a fashion consistent with Bayesian inference in one stimulus regime (e.g., ν_1) does not necessarily predict that the priors used under that stimulus regime would be the same as those under a different stimulus regime (e.g., ν_2):

$$\begin{aligned} \text{Posterior}(\nu_1) &\propto \text{Likelihood}(\nu_1) * \text{Prior}_1 \\ \text{Posterior}(\nu_2) &\propto \text{Likelihood}(\nu_2) * \text{Prior}_2. \end{aligned} \quad (1)$$

While the Prior probability in the Bayes rule has been interpreted as a probability distribution that reflects the statistics of the environment and hence, is stable and invariant to sensory conditions, this is not necessitated by the Bayes rule *per se*. Previously testing these questions have not been possible due to a lack of theoretical framework for fitting the priors, but which is now available (Körding et al., 2007). Here, we empirically tested whether priors are indeed stable in the face of substantial changes in the sensory conditions (see Equation 1). If we find that

subjects generate their posteriors under two different conditions (e.g. difference in visual contrast ν) according to Equation 1, then the likelihoods are expected to be different (i.e., $\text{Likelihood}(\nu_1) \neq \text{Likelihood}(\nu_2)$). On the other hand, the priors may or may not be different between the two conditions. The priors would only be the same (i.e., $\text{Prior}_1 = \text{Prior}_2$) if indeed they are independent of likelihoods.

To test this, we used an auditory-visual spatial localization task in which human performance was recently shown to be Bayes-optimal (Körding et al., 2007). Subjects were presented with visual stimulus as well as auditory stimulus and asked to report the perceived location of both stimuli. We tested each participant in two sessions one week apart. The two sessions were identical except for the contrast of the visual stimulus. The Bayesian model predicts a difference in the likelihood functions of the observers due to the contrast difference—which would in turn be reflected in a difference in posteriors and response. Considering that in our model the priors are estimated from observer responses, and that the observer responses are drastically different in the two sessions (due to the large change in visual contrast), it cannot be necessarily expected that the priors estimated from the two sessions would be equal. Therefore, finding equal priors in the two sessions would provide strong support for the proposition that the priors are encoded independently from the likelihoods.

We used spatial localization as the task for this study. Two features make this task of particular interest. First, it is a task in which the nervous system is implicitly engaged regularly in daily life. Although not always conscious of it, we constantly have to estimate the position of objects in order to navigate and interact with our environment and it is therefore a task which we expect to be optimized by evolution. Second, in some conditions there is a strong and well-known spatial localization illusion, the ventriloquist illusion, which illustrates the strong interactions between visual and auditory modalities in estimating the spatial location of objects. We presented observers with auditory and visual stimuli at variable locations and asked them to report their perceived locations of both visual and auditory stimuli. We scheduled the two sessions one week apart so that if there is any modulation of priors due to exposure to the first session, the effect would wear off by the time of the second session through exposure to the natural environment. Furthermore, a general assumption about priors is that they represent the statistics of the environment learned over the course of life or evolution, and are therefore, expected to be stable over time.

For each of the two sessions, we fitted the parameters of the likelihood and priors to the data and then compared the estimated likelihoods and priors between the two conditions. Part of the experimental data that we analyze here (for high contrast condition) has been reported previously (Körding et al., 2007).

Methods

Stimuli

Visual and auditory stimuli were presented independently at one of five positions. The five locations extended along a horizontal line 5° below the fixation point, from 10° to the left of the fixation point to 10° to the right of the fixation point, at 5° intervals. Visual stimuli were 35 ms presentations of Gabor wavelets extending 2° on a background of visual noise. The visual contrast was adjusted on an individual basis so that subjects' unimodal performance was 90% correct for the high contrast session and 40% correct for the low contrast session. Auditory stimuli were presented through a pair of headphones (Sennheiser HD280) and consisted of 35 ms white noise. The auditory stimuli were filtered through a Head Related Transfer Function (HRTF), gathered individually from subjects using a pair of in-ear microphones (Sound Professionals) using procedures similar to those described by <http://sound.media.mit.edu/KEMAR.html>, and simulated sounds originating from the five spatial locations in the frontoparallel plane where the visual stimuli were presented. In bimodal conditions, the auditory and visual stimuli were synchronized.

Procedure

Each observer participated in two sessions. The two sessions were identical in every way except for the contrast of the visual stimulus. In the first session the visual stimulus was high contrast and in the second session it was low-contrast. The second session was one week after the first session. Observers' task was to report the location of the visual stimulus as well as the location of the sound in each trial using the keyboard. The order of the auditory and visual reports was fixed within the experiment for each subject but was counterbalanced across subjects. Auditory and visual stimuli were presented alone or simultaneously, leading to a total of 35 conditions (see Figure 1). The experiment consisted of 15 trials of each condition, amounting to a total of 525 trials, ordered pseudo-randomly. Twenty naive observers (undergraduate and graduate students at Caltech, 18 to 35 years old, eleven male) participated in the experiment. The data for one participant was discarded since subsequent analysis showed that the auditory responses of the subject were at chance. Subjects were seated at a viewing distance of 54 cm from a 21-inch CRT monitor. A fixation cross was always present 5° above the level of stimuli, but its color turned from red to white 0.5 second before the start of the trial, and then remained that color throughout the trial. Participants were encouraged to take breaks every 10 minutes.

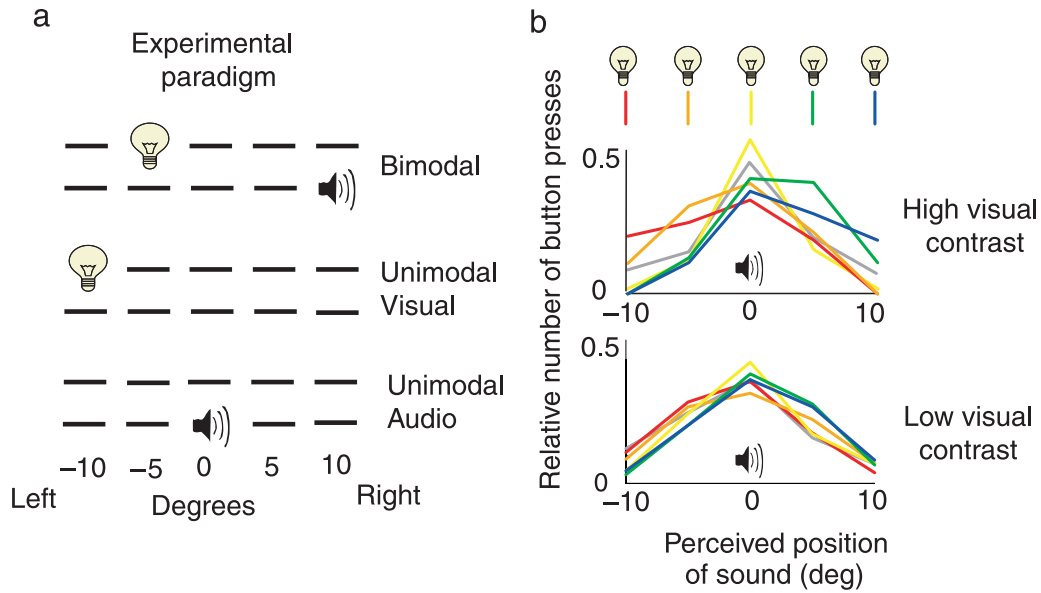


Figure 1. a) The experimental paradigm. Subjects were presented with either unimodal audio, unimodal visual or bimodal audio-visual stimuli. b) The influence of vision on the perceived position of an auditory stimulus in the central location is shown. Different colors correspond to the visual stimulus at different locations (sketched in warm to cold colors from the left to the right). The unimodal auditory case is shown in gray.

Model

We use the same model as that described in Körding et al. (2007). Figure 2 shows the statistical structure of the Bayesian observer model. The most important feature of this model is that it does not assume integration *a priori*. Instead it assumes that the sensory signals x_V and x_A are

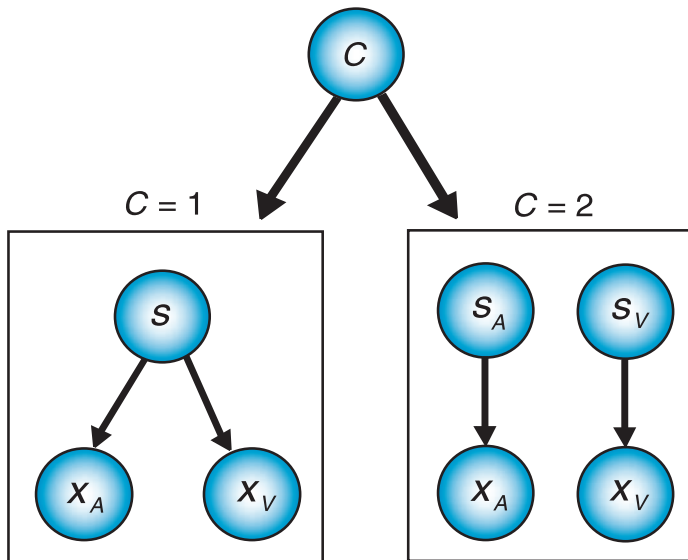


Figure 2. The Causal Inference model. The model assumes that each auditory and visual signal can be due to either one common cause ($C = 1$) or two independent causes ($C = 2$). Given the sensory signals x_V and x_A the brain has to infer which model is more likely and base its estimate of the sources s_V and s_A on this.

caused by either a single source s (Figure 2 left) or by two separate sources, s_A and s_V (Figure 2 right). x_V and x_A represent the visual and auditory signals, respectively, and are assumed to be conditionally independent, based on the observation that the auditory and visual signals are processed in separate pathways and are likely corrupted by independent noise.

Presented with the signals x_V and x_A the Bayesian observer therefore has to estimate whether the two signals originate from a common cause ($C = 1$) or from two separate causes ($C = 2$). How likely each scenario is depends on how similar the auditory and visual sensations (x_V and x_A) are. According to Bayes’ rule, the probability of there being a single cause is:

$$p(C = 1|x_V, x_A) = \frac{p(x_V, x_A|C = 1)p_c}{p(x_V, x_A|C = 1)p_c + p(x_V, x_A|C = 2)(1-p_c)}, \tag{2}$$

where $p_{c=}$ denotes the prior probability of a single cause in the environment and $p(x_V, x_A|C = 1)$ and $p(x_V, x_A|C = 2)$ can be found by marginalizing over s_A and s_V (see Körding et al., 2007). Given this knowledge, the optimal solution for the location that minimizes the mean expected squared error is:

$$\hat{s}_V = p(C = 1|x_V, x_A)\hat{s}_{C=1} + (1 - p(C = 1|x_V, x_A))\hat{s}_{V,C=2}, \tag{3}$$

$$\hat{s}_A = p(C = 1|x_V, x_A)\hat{s}_{C=1} + (1 - p(C = 1|x_V, x_A))\hat{s}_{A,C=2}, \tag{4}$$

where $\hat{s}_{V \text{ or } A}$ is the visual or audio response, $\hat{s}_{C=1}$ is the optimal estimate if we were certain that there is a single cause, and $\hat{s}_{V,C=2}$, $\hat{s}_{A,C=2}$ are visual and auditory uni-modal estimates, respectively, if we were certain that the two stimuli are independent (two causes). We assume that the unimodal likelihoods, $p(x_V|s_V)$, $p(x_A|s_A)$, as well as the prior probability distribution over locations (assuming $p(s) = p(s_V) = p(s_A)$), are normally distributed with means and variances (μ_A, σ_A^2) , (μ_V, σ_V^2) , and (μ_P, σ_P^2) , respectively. Thus:

$$\hat{s}_{V,C=1} = \hat{s}_{A,C=1} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_A}{\sigma_A^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}}, \quad (5)$$

and

$$\hat{s}_{V,C=2} = \frac{\frac{x_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}, \quad \hat{s}_{A,C=2} = \frac{\frac{x_A}{\sigma_A^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}}. \quad (6)$$

C is binomially distributed with $P(C = 1) = p_C$. We assume that the mean of the likelihoods are at the veridical locations and that mean of the prior distribution over locations is at the fixation point, 0 deg. In order to relate the theoretical posterior with the subjects' responses we assume that subjects try to limit their mean deviation and therefore report the mean of their posterior. The four free parameters (σ_A , σ_V , σ_P , p_C) were fitted to the participants' responses using 10000 trials of Monte Carlo simulation and MATLAB's *fminsearch* function (Mathworks, 2006), maximizing the likelihood of the parameters of the model.

The posterior can be rewritten in a more familiar form (Shams et al., 2005):

$$p(s_V, s_A | x_V, x_A) = \frac{p(x_V|s_V)p(x_A|s_A)p(s_V, s_A)}{p(x_V, x_A)}, \quad (7)$$

where

$$p(s_V, s_A) = p_c \delta(s_V - s_A) p(s_A) + (1 - p_c) p(s_V) p(s_A). \quad (8)$$

This is a mixture model (see Körding et al., 2007 for more details), mixing the prior from the two separate causal structures in Figure 2 and is therefore very similar to models developed for mixture problems for the visual system (Knill, 2003, 2007; Landy, Maloney, Johnston, & Young, 1995).

As in Körding et al. (2007) and Stocker and Simoncelli (2006), we model the trial to trial variability in observer responses as opposed to average behavior. We assume that the variability in response for the same stimulus condition

from trial to trial is primarily due to the noise in measurement (neuronal firing). Because of the variability in measurement, the mean of likelihood function fluctuates from trial to trial, but the variance is constant (here it is assumed that the nervous system has an accurate estimation of this variability). The average of the means of the likelihood distribution is assumed to be at the veridical position, i.e., no bias in measurement. The variability in the likelihood function leads to variability in posterior and the variability in the estimate (which is the mean of the posterior) from trial to trial.

Results

Figure 1b shows the subjects' auditory responses for various visual stimulus locations, but fixed auditory location. For visual stimuli presented to the left of the auditory stimulus, the subjects' responses naturally tend to shift to the left, and similarly to the right for visual stimuli presented to the right of the auditory response. The shift tends to be larger for high visual contrast than low visual contrast, as would be expected from any Bayesian model of multisensory interaction (Alais & Burr, 2004; Ernst & Banks, 2002; Ghahramani, 1995; Knill & Pouget, 2004; Körding et al., 2007; Shams et al., 2005).

To test the predictions of the model we first fitted the parameters of the causal inference model (Körding et al., 2007) to the data. The response probabilities of a representative human observer and the model for the high contrast data set are shown in Figure 3, where each panel corresponds to one stimulus condition. We calculated the goodness of fit R^2 over 300 data points (12 (\hat{s}_A, \hat{s}_V) combinations at 25 bimodal conditions). The average human observers' performance (pooled across subjects) is remarkably consistent with the Bayesian observer in the high contrast session, yielding $R^2 = 0.97$. The goodness of fit is also good, however lower, for the low contrast session, $R^2 = 0.75$, due to the larger variability in the visual data. The consistency of the human and Bayesian observer indicates that human sensory cue combination/segregation is Bayes-optimal. We have previously compared the performance of several different models on this task and found that the Causal Inference model performs the best among these (Körding et al., 2007).

Independence of priors from likelihoods

We examined the effect of change in the visual stimulus on the estimated likelihoods and priors. The likelihoods $p(x_A|s_A)$ and $p(x_V|s_V)$ are functions of the input, whereas the prior probabilities (p_C and $p(s)$) are generally assumed to be independent of the stimuli. Here we assume that the possible change in priors due to exposure to the uniform

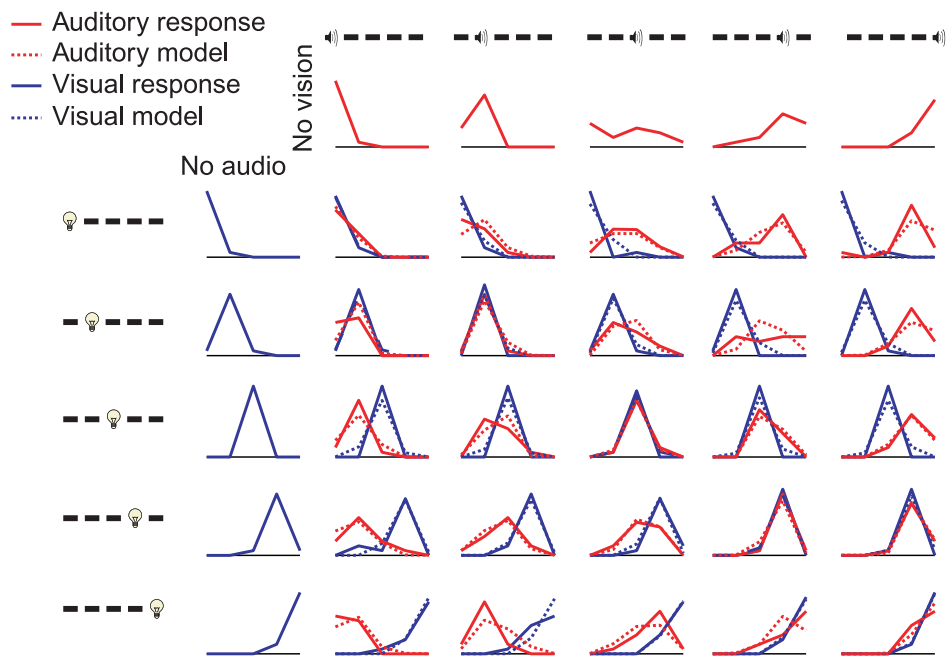


Figure 3. The 35 experimental conditions for one subject for high contrast session. Rows indicate visual, columns auditory conditions. Blue lines are the frequency of subject responses to the auditory stimuli, red lines indicate the frequency of visual responses to the visual stimuli and the dotted lines indicate the corresponding model fits.

distribution of stimuli in the first session is very small due to the short duration of the session (40 minutes), and this change, if any, decays after one week of exposure to normal environment, leaving the priors in effect unchanged. Given that the auditory stimulus was the same between the two sessions, we expected the auditory likelihood $p(x_A|s_A)$ to be the same across the two sessions. On the other hand, since the contrast of the visual stimulus was very different between the two sessions, we expected a noisier representation for the visual stimulus and thus, a broader likelihood distribution $p(x_V|s_V)$ in the low-contrast session. A change in the stimulus supposedly has no bearing on the prior knowledge about the environment, and thus the parameters characterizing the priors (p_C and σ_p) were expected to be the same between the two sessions. Indeed, the change in visual stimulus contrast led to considerable change in the visual performance; observers' performance in the visual-alone conditions declined on average by 40% in the low-contrast session. In contrast the average auditory performance declined only 2% in the low-contrast session.

We examined how the estimated likelihoods and priors differed between the two sessions, using multiple methods. We compared the parameters that were optimized separately for each session with each other to assess the difference in various parameters. As expected, the width of the visual likelihood is much larger (i.e., the precision is lower) in the low-contrast condition than in the high-contrast condition. As evidenced by the change in performance, this is a substantial change in the standard deviation of the likelihood distribution, from 2.12° to

11.71° . In contrast, the difference between the widths of the auditory likelihoods in the two conditions is minute (8.76° vs. 7.95°). This is consistent with the fact that the auditory stimulus was identical between the two sessions. These results confirm that these parameters indeed represent the theoretical notion of likelihood function and are not some arbitrary free parameters optimized to fit the data.

Next, we examined the change in the two prior parameters: the prior probability of a single cause and the width of the prior distribution over space. The change in these parameters is relatively small: p_C changes from 0.24 to 0.25 and the spatial prior variance, σ_p , changes from 11.55° to 13.12° . If the priors are the same in both sessions, using priors that are estimated from one data set should work as well as using priors that are optimized on the other data set. We tested this. Applying priors optimized from the high contrast data set to account for the low contrast data resulted in but a slight decrease in goodness of fit (from $R^2 = 0.75$ to $R^2 = 0.74$). Similarly, applying priors optimized from the low-contrast data to account for the high-contrast data resulted in only a slight decline in performance (from $R^2 = 0.97$ to $R^2 = 0.95$). Therefore, using priors optimized from a different data set caused hardly any decrease in goodness of fit.

These results suggest that the priors are approximately the same between the two sessions. While the difference between the prior parameters in two sessions is small, there is nevertheless some difference, and this small difference may reflect a meaningful effect. To examine this possibility, we fitted the parameters for the data sets

	Likelihoods		Priors	
	Visual σ_V	Auditory σ_A	Location σ_P	Common cause p_C
High (group)	2.12°	8.76°	11.55°	0.24
Low (group)	11.71°	7.95°	13.12°	0.25
High (Indiv)	2.1 ± 0.2°	9.2 ± 1.1°	12.3 ± 1.1°	0.28 ± 0.05
Low (Indiv)	15.0 ± 2.1°	9.4 ± 1.6°	15.8 ± 2.3°	0.24 ± 0.05

Table 1. Comparison of values of different parameters. For the individual data fitting we give mean ± standard error.

of individual participants (see Table 1) and for each parameter, we compared the value for the two sessions across participants using a two-tailed paired t -test (see Figure 4). The only parameter that showed a statistically significant difference between the two sessions is that associated with the visual likelihood (visual standard deviation, σ_V) ($p < 0.0005$). No other parameter had significantly different values across the two sessions ($p > 0.05$). Equivalently, the probability of replication is below 0.69 for each parameter except for σ_V (Prep > 0.995 , $z = 2.65$).

We have here assumed that the mean of the priors is fixed at 0° and that the mean of the likelihood is unbiased for all subjects. However, removing these constraints (by adding μ_P , and a bias to the likelihoods, as free parameters) does not change the results of the statistical tests above. Furthermore, neither one of these parameters undergoes a statistically significant change between the two sessions and the distribution of both parameters is not different ($p > 0.05$) from the assumed values (i.e., zero for

prior, and the veridical location for the likelihoods) in either session.

Altogether, these results suggest that, despite a large difference in the visual likelihood, the priors are stable across the two conditions. It should be noted that here we fail to reject the null hypothesis that the priors are equal between the two sessions. As this failure could in principle be due to an insufficient experimental power, we performed a statistical power analysis for the paired t -tests. The power to detect a low-mid size effect (a 0.5 standard deviation shift in the distribution) is moderately good (58%, 57% and 55% for σ_A , σ_P , and p_C , respectively). The power to detect a relatively large effect size (1 standard deviation) is excellent (99% for all three). Therefore, we can be highly confident that the change in the stimuli did not cause a large change in any of these three parameters, and can be fairly confident that it did not cause a moderate change. Therefore, the magnitude of difference would have to be quite small, if any, not to be detected by these tests. In light of the fact that the difference in visual likelihoods

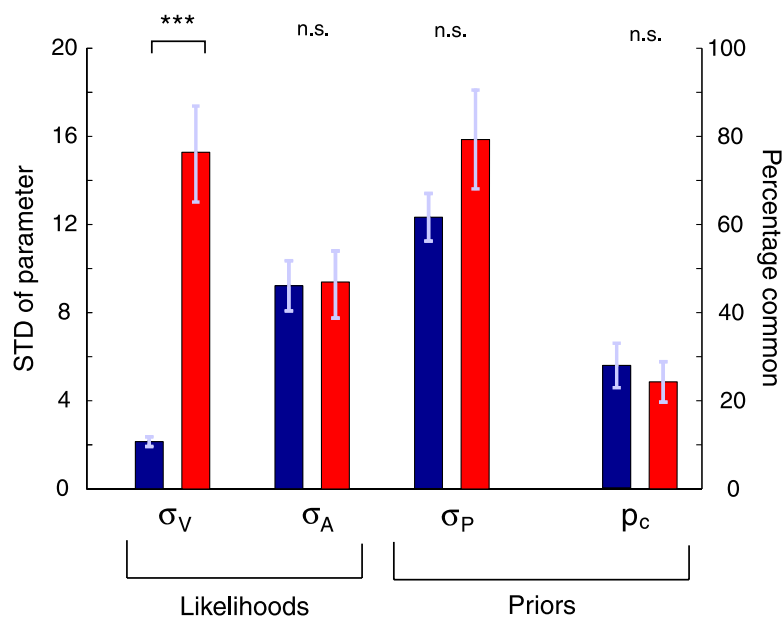


Figure 4. Estimated parameter values. Bar plot showing the value of the mean value (\pm standard error) of the four different parameters across subjects for the High contrast experiment (Blue) and Low contrast (Red) sessions separately. The only parameter that is significantly different between the two conditions was the standard deviation of visual likelihoods, σ_V , ($p < 0.0005$), no other parameter was different between the two sessions ($p > 0.05$).

is quite substantial (more than 10 standard deviations), such a putatively small change in priors would be negligible.

Discussion

Previously, we have reported that observers' responses in an auditory-visual spatial localization task is remarkably consistent with a Bayesian observer performing causal inference (Körding et al., 2007). Here, we show that drastically changing the stimuli has no statistically significant effect on the estimated prior probabilities. We found that using the priors estimated from a data set based on substantially different stimuli results in an excellent and equally good fit to the data as those estimated from the same data set. A direct comparison of the estimated parameters for the two conditions also confirmed that, while the likelihood function associated with the stimulus that was drastically changed is substantially affected by the change, the prior parameters remained fairly unchanged. Even if we had not changed the stimuli, i.e., if the two sessions were identical, the fact that the estimated priors are the same in two occasions that are one week apart, is note-worthy, as it suggests that priors are stable across time. Altogether, these results suggest that the priors are indeed independent of the stimuli and reflect *a priori* knowledge, and that the priors and likelihoods are represented independently in the nervous system and are combined according to Bayes rule in this perceptual task.

Recently, there has been much discussion of whether human behavior is Bayes-optimal (Rao, Olshausen, & Lewicki, 2002). For the behavior to be Bayes-optimal, the priors utilized in the inference process do not necessarily need to mirror the statistics of the environment. Even when the priors are stable, and independent of likelihoods, they may not reflect the true statistics of the environment if for some reason the observer has a wrong model of the world, or is constrained by other factors. Such an inference is nonetheless *subjectively* optimal even when the prior does not reflect the true statistics of the environment and is thus not *objectively* optimal. Here, we assumed that the observers have a prior bias for the center (straight ahead) location, and found that indeed this prior fits the data well and is stable across sensory conditions. If it is indeed true that most events fall in the straight-ahead location due to orienting behavior (observers quickly orient towards the events by an eye and head movement), then this prior could be considered *objectively* optimal. On the other hand, if this is not the case, and most auditory-visual events do not fall in the center of the auditory/visual field most of the time, then, this inference would only be *subjectively* optimal. Such a

prior might be due to evolutionary or biological constraints (i.e. 'hard-wired'); however, if the prior is modifiable by experience, then it is expected to reflect the true statistics of the environment as has been shown for the 'light-from-above' prior (Adams, Graf, & Ernst, 2004; Mamassian & Landy, 2001).

As we have shown, the framework of Bayesian inference provides experimenters with a principled approach to examining the role and nature of perceptual biases through evaluating Bayesian priors. Some work has already been done in this direction (e.g. Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson, 2002). Quantitatively estimating priors allows probing perceptual biases more concisely than previously possible by making it possible to explore the origins of these biases, whether the biases are stable across different conditions and observers, and whether these biases can be modified by experience, context, or other factors.

It is also worth mentioning the difference among the various types of perceptual priors so far studied in the literature. Stocker and Simoncelli (2006) presented subjects with moving stimuli under different visual contrasts and were able to estimate a prior on visual velocities. In contrast here we have studied a prior that is composed of two components, one over spatial location, $p(s)$, (the counterpart of the prior on velocities in Stocker and Simoncelli's study), as well as a component, p_C , that encapsulates the expected probability of two auditory and visual sources being the same, and hence specifies the degree of interaction between two modalities, similar to Bresciani et al. (2006) and Shams et al. (2005). Although the prior was constant across conditions in this study, it is expected that it would vary for different tasks and different modalities. While the *a priori* expectation of a common cause is expected to be mostly due to the learned or hard-wired statistics of the auditory-visual events in the environment, it may also be affected by the instructions provided to the observer by the experimenter or the context of the experiment (Ernst, 2007). It seems highly likely that some of the differences in the crossmodal interactions reported by different studies are due to differences in the prior expectation of the common cause, p_C (see Hospedales & Vijayakumar, 2009 for a recent analysis).

These findings, together with many other recent findings (Alais & Burr, 2004; Battaglia et al., 2003; Bühlhoff & Mallot, 1988; Ernst & Banks, 2002; Jacobs, 1999; Körding et al., 2007; Körding & Tenenbaum, 2007; Shams et al., 2005; Stocker & Simoncelli, 2006; van Beers et al., 1999), provide accumulating evidence that the nervous system utilizes Bayesian inference. However, they do not describe how the probability distributions and operations required for this computation are implemented in the brain. Recent work on this question has led to promising results, e.g. Ma, Beck, Latham, and Pouget (2006) have shown how in theory the multiplication of

likelihood and prior is a natural outcome of population coding with biologically realistic neurons and Poisson distributed firing rates.

Conclusions

We have shown that the priors in an audio-visual localization task are independent of likelihoods and are thus encoded separately. This finding is consistent with the general notion that the nervous system combines sensory estimates with the prior knowledge about the environment for perceptual processing.

Acknowledgments

We thank Konrad Koerding, Wei Ji Ma, Stefan Schaal and Peter Bossaerts for their insightful discussions and comments. We also wish to thank the anonymous reviewers for some very useful comments. U.B. and S.Q. were supported by the David and Lucile Packard Foundation and the Betty and Gordon Moore Foundation. L.S. was supported by UCLA Faculty Grants Program and Career Development Program.

Commercial relationships: none.

Corresponding author: Ladan Shams.

Email: ladan@psych.ucla.edu.

Address: UCLA Psychology Department, Los Angeles, CA 90095, USA.

References

- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the ‘light-from-above’ prior. *Nature Neuroscience*, *7*, 1057–1058. [[PubMed](#)]
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. [[PubMed](#)] [[Article](#)]
- 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, and Vision*, *20*, 1391–1397. [[PubMed](#)]
- Beierholm, U., Kording, K., Shams, L., & Ma, W. J. (2008). Comparing Bayesian models for multisensory cue combination without mandatory integration. *Advances in neural information processing systems* (vol. 20, pp. 81–88). Cambridge, MA: MIT Press.
- 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://journalofvision.org/6/5/2/>, doi:10.1167/6.5.2. [[PubMed](#)] [[Article](#)]
- Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A, Optics and Image Science*, *5*, 1749–1758. [[PubMed](#)]
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5):7, 1–14, <http://journalofvision.org/7/5/7/>, doi:10.1167/7.5.7. [[PubMed](#)] [[Article](#)]
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [[PubMed](#)]
- Ghahramani, Z. (1995). *Computational and psychophysics of sensorimotor integration*. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology.
- Hospedales, T., & Vijayakumar, S. (2009). Multisensory oddity detection as Bayesian inference. *PLoS ONE*, *4*, e4205. [[PubMed](#)]
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621–3629. [[PubMed](#)]
- Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, *43*, 831–854. [[PubMed](#)]
- 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://journalofvision.org/7/7/5/>, doi:10.1167/7.7.5. [[PubMed](#)] [[Article](#)]
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*, 712–719. [[PubMed](#)]
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*, e943. [[PubMed](#)] [[Article](#)]
- Körding, K. P., & Tenenbaum, J. B. (2007). Causal inference in sensorimotor integration. *Advances in neural information processing systems* (vol. 19, pp. 737–744). Cambridge, MA: MIT Press.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412. [[PubMed](#)]
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438. [[PubMed](#)]

- Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research*, *41*, 2653–2668. [[PubMed](#)]
- Rao, R., Olshausen, B., & Lewicki, M. (2002). *Probabilistic models of the brain*. Cambridge, Massachusetts: MIT Press.
- 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 of the Royal Society of London B: Biological Sciences*, *273*, 2159–2168. [[PubMed](#)] [[Article](#)]
- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*, 1923–1927. [[PubMed](#)]
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*, 578–585. [[PubMed](#)]
- van Beers, R. J., Sittig, A. C., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*, 1355–1364. [[PubMed](#)] [[Article](#)]
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604. [[PubMed](#)]