

# Causal inference in multisensory perception

## Supporting Information

### Comparison with other models

In recent models of cue combination [1-3] the response distribution  $p(\hat{s}_V, \hat{s}_A | s_V, s_A)$  is treated as the posterior distribution  $p(s_V, s_A | x_V, x_A)$ . This treatment leads to a significant error for non-Gaussian priors. In this paper we correctly deal with this issue by marginalizing over the latent variables  $x_A$  and  $x_V$ :

$$p(\hat{s}_V, \hat{s}_A | s_V, s_A) = \int \int p(\hat{s}_V, \hat{s}_A | x_V, x_A) p(x_V | s_V) p(x_A | s_A) dx_V dx_A.$$

These response distributions were obtained through simulation.

The results reported for these other models here correspond to the corrected versions of these models.

### Priors of other models

Recently, models with so-called interaction priors have been proposed [1-3], which define explicit interactions between the cues. These models do not assume full integration and thus lead to much better predictions than the traditional forced-fusion model. While the model by Shams et al. [2] was the first of such models, we will not make a comparison with that model here as it has many more free parameters than the causal inference model (and in fact it may be seen as a superset of all the models tested here). Table 1 in the main text analyses how well each of the models predicts the data.

Given an interaction prior  $p(s_A, s_V)$ , we obtain the bisensory posterior distribution through Bayes' rule,

$$p(s_A, s_V | x_A, x_V) \propto p(s_A, s_V) p(x_A | s_A) p(x_V | s_V).$$

In earlier models (Bresciani et al.; Roach et al.), the interaction prior is assumed to take a specific form. In Bresciani et al., it is a Gaussian ridge on the diagonal with width

$\sigma_{\text{coupling}}$  :

$$p(s_A, s_V) \propto e^{-\frac{(s_A - s_V)^2}{2\sigma_{\text{coupling}}^2}}$$

A straightforward computation then allows us to obtain the auditory posterior:

$$\begin{aligned} p(s_A | x_A, x_V) &\propto \int p(s_A, s_V) p(x_A | s_A) p(x_V | s_V) ds_V \\ &\propto \int e^{-\frac{(s_A - s_V)^2}{2\sigma_{\text{coupling}}^2}} p(x_A | s_A) p(x_V | s_V) ds_V \\ &\propto p(x_A | s_A) \int N(s_V; s_A, \sigma_{\text{coupling}}) N(s_V; x_V, \sigma_V) ds_V \\ &= N(s_A; x_A, \sigma_A) N\left(s_A; x_V, \sqrt{\sigma_{\text{coupling}}^2 + \sigma_V^2}\right) \\ &\propto N\left(s_A; \frac{x_A \sigma_A^{-2} + x_V (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}{\sigma_A^{-2} + (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}, \frac{1}{\sqrt{\sigma_A^{-2} + (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}}\right) \end{aligned}$$

where  $N(x; \mu, \sigma)$  is the value of the normal distribution with mean  $\mu$  and standard deviation  $\sigma$  evaluated at  $x$ . Because the posterior distribution is a pure Gaussian, it can be identified with the response distribution. The expression for the visual distribution is obtained by interchanging  $V$  and  $A$ .

In the work by Roach et al. [1], a more general form of interaction prior is used, consisting of the same Gaussian ridge on the diagonal, but added to that a constant background  $\omega$ :

$$p(s_A, s_V) \propto \omega + e^{-\frac{(s_A - s_V)^2}{2\sigma_{\text{coupling}}^2}}$$

This gives rise to the following posterior distribution:

$$\begin{aligned} p(s_A | x_A, x_V) &\propto \int \left( \omega + e^{-\frac{(s_A - s_V)^2}{2\sigma_{\text{coupling}}^2}} \right) p(x_A | s_A) p(x_V | s_V) ds_V \\ &= \omega p(x_A | s_A) \int p(x_V | s_V) ds_V + p(x_A | s_A) \sqrt{2\pi\sigma_{\text{coupling}}^2} \int N(s_V; s_A, \sigma_{\text{coupling}}) N(s_V; x_V, \sigma_V) ds_V \\ &= \omega N(s_A; x_A, \sigma_A) + \sqrt{2\pi\sigma_{\text{coupling}}^2} N(s_A; x_A, \sigma_A) N(s_A; x_V, \sqrt{\sigma_{\text{coupling}}^2 + \sigma_V^2}) \\ &= \omega N(s_A; x_A, \sigma_A) + \\ &\quad + \sqrt{2\pi\sigma_{\text{coupling}}^2} N(x_A; x_V, \sqrt{\sigma_{\text{coupling}}^2 + \sigma_V^2 + \sigma_A^2}) N\left(s_A; \frac{x_A \sigma_A^{-2} + x_V (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}{\sigma_A^{-2} + (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}, \frac{1}{\sqrt{\sigma_A^{-2} + (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}}\right) \end{aligned}$$

The best estimate under is the mean of this distribution, which is

$$\hat{s}_A = \frac{\omega x_A + \sqrt{2\pi\sigma_{\text{coupling}}^2} N(x_A; x_V, \sqrt{\sigma_{\text{coupling}}^2 + \sigma_V^2 + \sigma_A^2}) \frac{x_A \sigma_A^{-2} + x_V (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}{\sigma_A^{-2} + (\sigma_{\text{coupling}}^2 + \sigma_V^2)^{-1}}}{\omega + \sqrt{2\pi\sigma_{\text{coupling}}^2} N(x_A; x_V, \sqrt{\sigma_{\text{coupling}}^2 + \sigma_V^2 + \sigma_A^2})}$$

The expression for the visual estimate is obtained by interchanging  $V$  and  $A$ . This model is more similar to the causal inference model, because the estimates are nonlinear functions of  $x_V$  and  $x_A$ . Because of that reason as well, the response distribution can now no longer be identified with the response distribution.

There is a direct link between the causal inference model and these models with interaction priors. The causal inference model can be recast as a model with an interaction prior by integrating out the latent variable  $C$ :

$$p(s_V, s_A) = p_{\text{common}} \delta(s_V - s_A) p(s_A) + (1 - p_{\text{common}}) p(s_V) p(s_A).$$

All predictions about position estimates with this model are retained. However, this model no longer explicitly represents whether there is a single cause or alternatively two independent causes. This explains why those models are relatively successful at explaining the data.

Fig 1S shows the priors obtained from the models discussed here, after fitting their parameters to optimally describe the data.

### **Bias as a function of disparity**

In the main text we have used the probability of the data given the model as a measure for the quality of each model. We found that the causal inference model best explains the data. Although such inferential statistics are a good tool to compare models, we are also interested in the question why the causal inference model performs better. To get an understanding of the differences we turn to the graph showing bias as a function of disparity (Fig. 2S). The data shows that the bias decreases with increasing spatial disparity. The further the distance between visual and auditory stimuli, the smaller is the influence of vision on audition. This result is naturally predicted by the causal inference model; larger discrepancies make the single cause model less likely as it needs to assume large noise values that are unlikely. The joint prior used by Bresciani et al. [3] predicts a bias that is largely invariant to the disparity. However, lacking a way to represent that two cues may be entirely independent, it underestimates the derivative of the bias graph.

### **References**

1. Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc Biol Sci* 273: 2159-2168.
2. Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. *Neuroreport* 16: 1923-1927.
3. Bresciani JP, Dammeier F, Ernst MO (2006) Vision and touch are automatically integrated for the perception of sequences of events. *J Vis* 6: 554-564.

## Figure Legends:

**Figure 1, supporting information:** The interaction priors when fit to our dataset are shown for the causal inference model, the Roach et al. [1] and the Bresciani et al. priors[3].

**Figure 2, supporting information:** The average auditory bias  $\frac{\hat{s}_A - s_A}{s_V - s_A}$ , i.e. the relative influence of the visual position on the perceived auditory position, is shown as a function of the absolute spatial disparity (solid line, as in Fig. 2 main text) along with the model predictions (dashed lines). Red: causal inference model. Green: behavior derived from using the Roach et al prior. Purple: behaviour derived from using the Bresciani et al prior.