

Evidence accumulation across the senses in the face of causal uncertainty

Jochem Beurskens*

Donders Centre for Neuroscience (DCN), Radboud University
Heyendaalseweg 135, Nijmegen, The Netherlands
jochem.beurskens@donders.ru.nl

Steffen Buegers*

Royal HaskoningDHV
Centre for Computational Neuroscience and Cognitive Robotics (CNCR), University of Birmingham
B15 2TT, Birmingham, UK
steffen.burgers@rhdhv.com

Agata Wlasczyk

Department of Applied Mathematics and Computer Science, Technical University of Denmark
Richard Petersens Plads 2800 Kgs. Lyngby, Denmark
agwl@dtu.dk

Uta Noppeney

Donders Centre for Neuroscience (DCN), Radboud University
Heyendaalseweg 135, Nijmegen, The Netherlands
Centre for Computational Neuroscience and Cognitive Robotics (CNCR), University of Birmingham
B15 2TT, Birmingham, UK
uta.noppeney@donders.ru.nl

Abstract

Humans need to make accurate and timely decisions based on a constant influx of noisy sensory signals from common and separate causes. Current Bayesian Causal Inference (BCI) models have shed light on how the brain arbitrates between sensory integration and segregation when uncertain about the signals' causal structure. Yet, these static models made predictions only for response choices and ignored the dynamic nature of perceptual decision making. Using psychophysics, we show that crossmodal biases (CMB) decline for longer response times. We then develop a dynamic Bayesian causal inference (BCI) model that accumulates evidence jointly about an event's location and the signals' causal structure over time until a heuristic decisional threshold is reached. This dynamic BCI model accounts for the decline in crossmodal influences over time by the progressive resolution of spatial and causal uncertainty.

Keywords: Bayesian causal inference; Evidence accumulation; Behavioural modelling; Ventriloquist effect;

Introduction

In our natural environment observers need to make accurate and timely decisions based on a constant influx of noisy signals. A wealth of work in unisensory perception has shown that observers accumulate multiple sensory samples until a decisional threshold is reached consistent with normative sequential sampling models (Gold & Shadlen, 2007). Further, observers have been shown to integrate sensory signals from common causes weighted by their momentary reliabilities and accumulate this integrated evidence over time (Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014). However, in natural situations observers do not a priori know whether signals come from common sources and should hence be integrated. They need to infer the signals' causal structure from noisy cross-sensory correspondence cues such as signals happening at the same time or location (Noppeney, 2021). In the face of this causal uncertainty observers need to concurrently accumulate evidence about environmental properties such as an object's location and the signals' causal structure leading to complex non-linearities in the decision process.

This study combines psychophysics in human observers and Bayesian modelling to investigate how the brain accumulates evidence about an event's location from auditory and visual senses under causal uncertainty into timely and accurate perceptual decisions.

Methods

Experimental procedure

In a spatial ventriloquist paradigm 15 observers were presented with synchronous, spatially congruent and disparate audiovisual signals. On each trial, auditory (A) and visual (V) locations were independently sampled from 8 equally spaced locations, ranging from -21° to 21° along the azimuth. In separate blocks observers reported either their perceived A or V

location by pressing one of 8 buttons under speed or accuracy instructions.

Generative model

The generative model (figure 1A) assumes that common and independent cause cases ($C=1$ or $C=2$) are sampled from a binomial distribution as defined by a causal prior p_{common} (Körding et al., 2007). A and V stimulus locations (sA, sV) are sampled jointly ($C=1$) or independently ($C=2$) from a spatial prior distribution $N(\mu_p, \sigma_p)$. On each trial, these stimulus locations generate a series of independent noisy sensory observations ($xA_{\leq t}, xV_{\leq t}$).

Recognition model

Given a series of noisy observations (xA_t, xV_t), the observer is assumed to compute the posterior distribution over the causal structure C and the A location (similarly for V location):

$$P(sA, C | xA_{\leq t}, xV_{\leq t}) = \frac{P(xA_t, xV_t | sA, C) P(sA, C | xA_{\leq t-1}, xV_{\leq t-1})}{\sum_C \int P(xA_t, xV_t | sA, C) P(sA, C | xA_{\leq t-1}, xV_{\leq t-1}) dsA}$$

for $t = 1$:

$$P(sA | C, xA_{\leq t-1}, xV_{\leq t-1}) = P(sA | C) \text{ (i.e. spatial prior)}$$

$$P(C | xA_{\leq t-1}, xV_{\leq t-1}) = P(C = 1) \text{ (i.e. causal prior)}$$

We obtain the posterior probability over A (or V) location by marginalizing over the unknown causal structure C:

$$P(sA | xA_{\leq t}, xV_{\leq t}) = P(sA, C = 1 | xA_{\leq t}, xV_{\leq t}) + P(sA, C = 2 | xA_{\leq t}, xV_{\leq t})$$

When the observer responds, they read out the final A or V location estimate (\hat{sA} or \hat{sV}) as either the mean or the MAP of the evolving posterior distribution (figure 1B). The continuous location read outs are mapped onto the closest button responses for comparison with observers' discrete response data. In addition, we explored the mapping from the model's discrete time sample to observers' response times by manipulating the sample duration (2500 ms, 625 ms, 312.5 ms, 156.25 ms, 78.125 ms).

At each time point, the observer decides whether to stop/respond or continue sampling based on this evolving posterior distribution. We assessed the following stopping criteria for the MAP estimate, and criteria iii and iv for the mean estimate: i) $P(sA = \hat{sA} | xA_{\leq t}, xV_{\leq t}) > crit$, ii) $P(sA = \hat{sA} | xA_{\leq t}, xV_{\leq t}) - P(sA = \hat{sA} | xA_{\leq t-1}, xV_{\leq t-1}) \leq crit$, iii) normalized Shannon entropy $E_N \geq crit$ (Li & Ma, 2020), iv) $E_N(t) - E_N(t-1) \leq crit$.

We fitted the 6 (stopping criterion & read out) x 5 (sample duration) models to individual's joint response time (RT, 64 time bins of 39.0625 ms duration) x choice (8 buttons) histograms, by minimising the negative log likelihood of the model responses given the data. At the random effects (i.e. group level), we obtain the best model with respect to: i. sample duration and ii. stopping criterion & read out, using factorial model comparison (Acerbi, Dokka, Angelaki, & Ma, 2018).

Results

Behaviour

The crossmodal bias (CMB) (e.g. for A report: $(Areport_{sA=x, sV=y} - Areport_{sA=sV=y})$) quantifies the in-

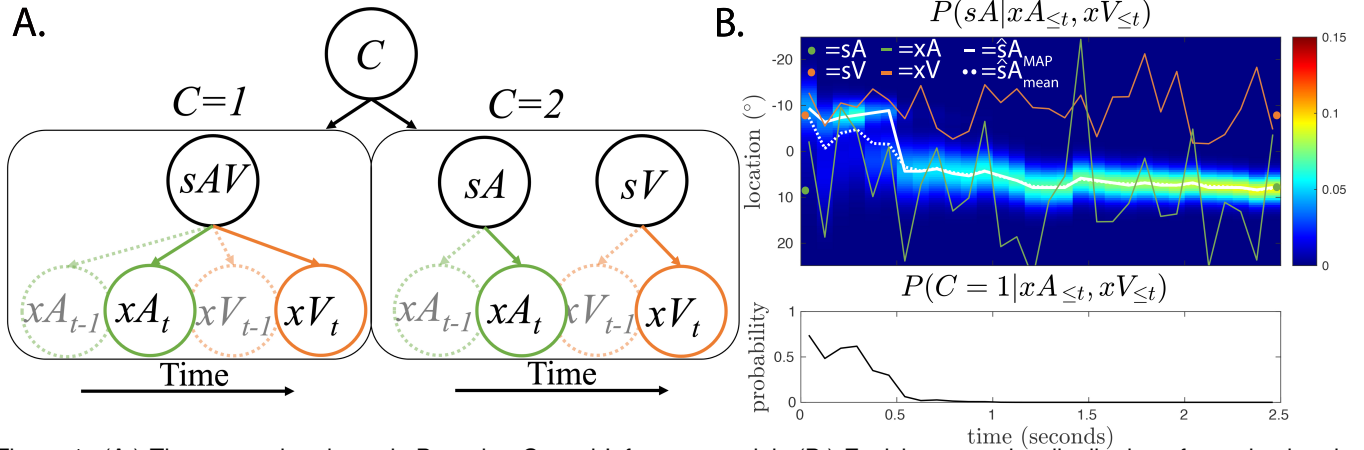


Figure 1: (A.) The generative dynamic Bayesian Causal Inference model. (B.) Evolving posterior distributions for a simulated trial with $sA = 9$ and $sV = -9$. Based on a series of noisy observations, xA_t and xV_t the posteriors, and the MAP and mean estimates evolve over time.

fluence of V location on observers' perceived sound location. Consistent with BCI the crossmodal bias initially increases, but then decreases again for greater spatial disparities. This makes common causes for A and V signals progressively unlikely. Importantly, the non-linear CMB function depends on observers' RTs. On fast trials (RTs within 1st tercile for each participant), the CMB profile is more similar to sensory fusion with overall greater CMB and less CMB decline for larger spatial disparities (figure 2).

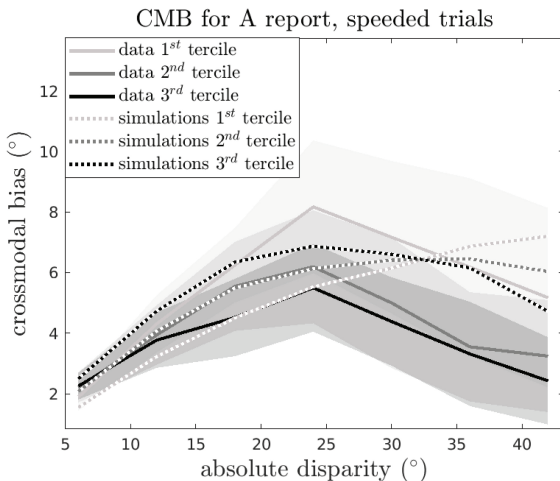


Figure 2: Crossmodal biases for observers' responses (across participants' tertiles mean \pm 95% CI) and model simulation (stopping criterium $E_N(t) - E_N(t - 1)$ and readout MAP). The data is split into tertiles based on the response times.

Modelling

Qualitative: The dynamic BCI model accounts for observers' CMB profile across RT and disparity by spatial uncertainty principles. Initially, the model faces large spatial and hence causal uncertainty, so that the spatial estimate resembles a vision-dominated fusion estimate. With increasing number of

samples this spatial uncertainty gradually resolves resulting in less crossmodal influences (figure 1B).

Quantitative: Factorial model comparison at the group level reveals superior performance of models that map samples to a duration of 0.3125 ms (protected exceedance probability $P_{exp} = 0.636$). Models with heuristic stopping criteria of $E_N(t) - E_N(t - 1) \leq crit$ and a MAP estimate account better for observers' response profile ($P_{exp} = 0.304$). However, there was considerable inter-observer variability with respect to the sample duration, the stopping criteria and read outs, which will be explored in further communication.

Discussion

Previous static Bayesian Causal Inference models explained how the brain arbitrates between sensory integration and segregation depending on the world's causal structure (Körding et al., 2007). However, they ignored the dynamics of perceptual decision making allowing predictions only for response choices but not for response times. This dynamic Bayesian causal inference model sheds light on why observers' response choices and crossmodal biases depend on their response times. Evidence accumulation progressively resolves observers' spatial uncertainty about the A and V locations, and thereby their uncertainty about the signals' causal structure. This decline in causal uncertainty in turn reduces the influence of task-irrelevant spatially disparate inputs from other sensory modalities. Moreover, this same computational mechanism can also explain recent neuroimaging findings suggesting a putative progression from earlier sensory fusion estimates in posterior parietal cortices to Bayesian Causal inference estimates in anterior parietal cortices (Aller & Noppeney, 2019; Rohe, Ehrlis, & Noppeney, 2019; Cao, Summerfield, Park, Giordano, & Kayser, 2019). Future research will explore the considerable inter-observer variability in heuristic stopping criteria.

References

- Acerbi, L., Dokka, K., Angelaki, D. E., & Ma, W. J. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS computational biology*, *14*(7), e1006110.
- Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical bayesian causal inference. *PLoS biology*, *17*(4), e3000210.
- Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron*, *102*(5), 1076–1087.
- Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *Elife*, *3*, e03005.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.
- Körding, 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.
- Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the bayesian confidence hypothesis. *Nature communications*, *11*(1), 2004.
- Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual review of neuroscience*, *44*, 449–473.
- Rohe, T., Ehlis, A.-C., & Noppeney, U. (2019). The neural dynamics of hierarchical bayesian causal inference in multisensory perception. *Nature communications*, *10*(1), 1907.