Causal inference can explain hierarchical motion perception and is reflected in neural responses in MT

Causal inference (CI) has recently been proposed as a universal computational motif in the brain [Shams & Beierholm 2020]. However, how CI is implemented by neural circuits, and its signatures in terms of single neuron responses, are still unclear. We have investigated this question in the context of complex motion processing. Motion perception deviates from retinal motion [Johansson 1973] a computation that can be understood in terms of hierarchical CI over which moving elements to integrate into coherent 'groups' vs segment into different ones [Gershman et al. 2016, Shivkumar et al. 2020]. Yet, most of our understanding of the neural basis of motion processing is in terms of retinal motion, delegating potential CI computations to downstream cortical areas [Rohe et al. 2015, 2019].

Our work makes two contributions: first, we present new psychophysical evidence for the hierarchical nature of this process using a display of hierarchically nested groups of moving dots. Second, we use the hierarchical CI model fit to psychophysical data to derive quantitative neural predictions for neurons representing the variables in our model. At each level, our model contains two types of variables: one that represents the retinal motion predicted by the larger surround, and one that represents the difference between the actual local motion and that predicted from the surround. The predicted neural responses show remarkable similarity to two classes of neurons found in area MT: neurons with suppressing and with reinforcing surrounds [Born & Bradley 2005]. Finally, we present new neurophysiological data from area MT in a macaque monkey where the velocity-dependent pattern of surround suppression of neural responses agreed with that predicted for the relative variable in our CI model. Our results show that signatures of CI are already present at the early stages of sensory processing, and suggest that they may be implemented by local computations.

Our work builds on a recent model that modified and reformulated the model by Gershman et al. [2016] as hierarchical causal inference based on local computations [Shivkumar et al. 2020]. The key elements of this model (Fig. 1A) are a decomposition of the motion at every location ('dot') into motion that is predicted by a larger entity ('group') that the motion at that location is inferred to belong to, plus a relative motion: $\vec{v}^{\text{dot}}_{\text{retina}} = \vec{v}^{\text{group}}_{\text{retina}} + \vec{v}^{\text{dot}}_{\text{relative}}$. This process is repeated hierarchically with groups being inferred to be part of 'supergroups' etc. A key innovation of this model is the mixture prior consisting of a delta around zero plus a slow speed prior [Stocker and Simoncelli 2006] that ensures that the model only infers non-zero relative motion if the local motion sufficiently deviates from the predicted motion. As a result of this prior, the model 'chunks' the visual scene into a hierarchical structure consisting of parts that are inferred based on their coherent motion (each consisting of elements with zero relative motion). Since human motion perception has been shown to be relative to the next-larger entity a visual element belongs to, it corresponds to the non-zero relative motion variable in this model that is lowest in the hierarchy. E.g. if $\vec{v}^{\text{dot},k}_{\text{retina}} = 0$ for some k, then this dot is perceived to move at the velocity of the group that it belongs to, $\vec{v}^{\text{group},n}_{\text{relative}}$, if that is non-zero, otherwise the relative velocity at the next higher level etc.

We psychophysically tested the three key predictions of this model: integration of similarly moving elements with an inferred motion that is based on cue-combination, segmentation of differently moving elements leading to the perception of relative motion [Johanssen 1973], and percepts of increased uncertainty in between. Experiment 1 (Fig 1B) consisting of only red and green dots found clear evidence for both segmentation (perception of relative motion) and a bimodal distribution of responses in the transition region between integration and segmentation. However, since integration in this case implies only a minimal change in percept compared to retinal motion, we designed a 2nd experiment with an additional, larger surround (blue dots in Fig. 1B). Now, whenever the motion of the green dots is integrated with that of the red dots, they are predicted to be perceived relative to the blue dots, implying (by experimental design) a deviation in the opposite direction from that predicted when the green dots are inferred to be moving differently from the red dots. The data from this experiment (Fig 1D) now clearly demonstrates both integration and segmentation, allowing us to quantitatively test our CI model and infer its parameters.

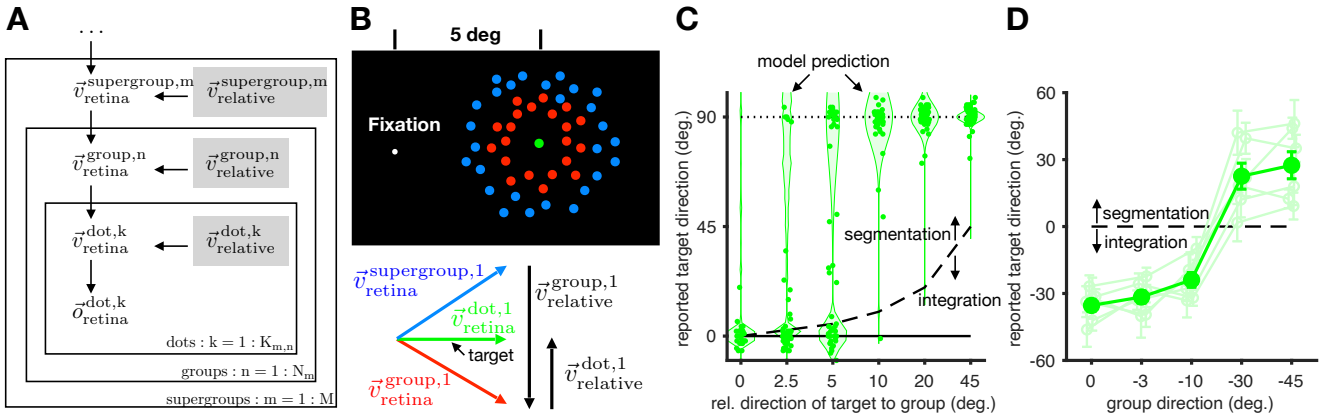

Figure 1. A: Generative model. **B:** Stimuli with corresponding velocity vectors. **C:** Example observer reporting perceived motion direction of green dot. **D:** Responses of 6 observers (thin) plus average (thick line).

Our computational model allows us to generate predictions for the responses of neurons representing the individual latent variables in our model. Here we focus on the lowest hierarchical level in our model and compare to neural recordings in area MT. We map the posteriors computed in our model to neural responses in the following way: we start by measuring the speed tuning and motion direction tuning of each MT neuron for a simple stimulus consisting only of one the central motion element (green in Fig 1B), with a stationary surround (red in Fig 1B; both white on black for the neurophysiological experiments). This yields the neuron's tuning to the posterior over the underlying variable regardless of whether it represents a retinal or relative variable in our model since the posteriors over each variables are identical for a stationary surround. However, as soon as we introduce motion into the surround, the posteriors over the retinal and relative variables in our model diverge sharply in their mean. As long as there is little uncertainty over the causal structure, the respective posteriors have approximately the same means and variances as those encountered when measuring the speed and direction tuning with a stationary surround, allowing us to look up the corresponding neural response. When center and surround velocities are similar and the implied posterior is bimodal, we compute the neural prediction as a combination of both mixture components.

Fig. 2A+B show the neural predictions for different combinations of center and surround directions assuming both to have the same speed, and assuming a linearly increasing speed tuning for the neuron. While the retinal neuron's response is largely independent of the surround direction (with the exception of the integration zone), the relative neuron's response is highly sensitive to center and surround directions. Focusing on the surround-suppressed neurons in a newly collected dataset, we find a good agreement between the model predictions and data when comparing tuning curves for stationary surround with a surround moving at −90deg relative to the preferred direction of the neuron (Fig. 2C+B). Since each neuron has somewhat different speed and direction tuning, we only show a model fit for an example neuron (Fig. 2C+D). But the qualitative pattern holds across our population (Fig. 2E).

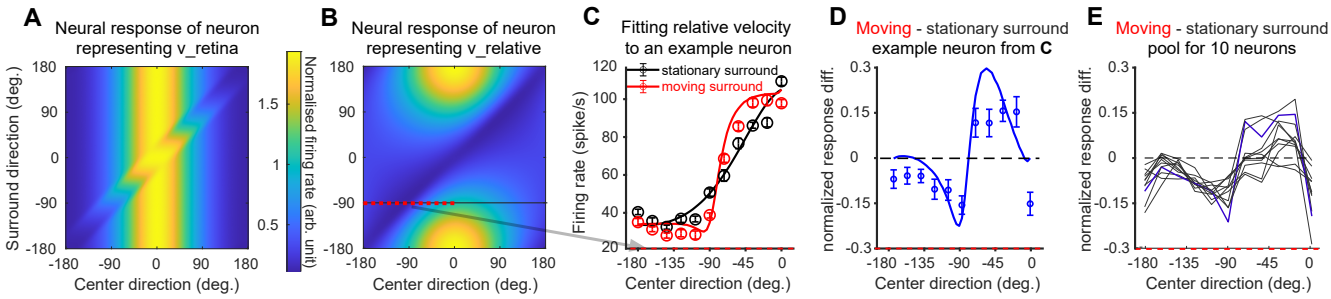

Figure 2. A & B: Neural prediction for neuron representing $\vec{v}^{\text{dot}}_{\text{retina}}$ (A) and $\vec{v}^{\text{dot}}_{\text{relative}}$ (B). **C:** Example unit with moving (red error bars) and stationary surround (black error bars) as a function of center motion direction with surround motion at −90°. Lines are model fits. **D:** The normalized response difference between moving and stationary surround (blue error bars) for the example unit in (C) along with the CI model's prediction (blue line). **E:** Same as (C), but showing the individual response differences for 10 neurons (gray lines). Blue line indicates the example neuron from C & D. Individual lines not expected to be identical since neurons differ somewhat in direction and speed tuning.