Human visual motion perception shows hallmarks of Bayesian structural inference

SUMMARY

Bayesian inference has emerged as a successful description of elementary human visual motion perception, but little is known about how we make sense of the complex, nested motion relations found in real-world dynamic scenes. Here we expand the computational understanding of human motion perception to the domain of structural inference by virtue of two theory-driven experiments. To do so, we leveraged a recently proposed framework for generating analytically tractable motion-structured stimuli. A first experiment asked participants to categorize the (often hierarchically nested) motion structure of short dynamic scenes. Their psychometric function, when analyzed in terms of the statistically correct Bayesian posterior—a key facility of our tractable stimulus design—, revealed the signature logistic shape of correct structural inference. We noticed that participants exhibited distinct perceptual error patterns between certain structures. A Bayesian observer model linking human responses to the statistical evidence in individual trials was able to account for these patterns and, even more, explained participants' responses with single trial resolution. We therefore hypothesized that the Bayesian description could provide a general account for human motion structure perception.

We tested this hypothesis in a second experiment: a two-alternative forced choice task targeting human perception of highly ambiguous scenes, which we generated by morphing continuously between two prototypical motion structures. Strikingly, the Bayesian observer model from the first experiment qualitatively predicted human choices for the new, ambiguous scenes, thereby confirming the Bayesian model's generality. As an additional hallmark of probabilistic reasoning, participants' reported decision confidence correlated with the certainty in the correct posterior distribution.

Taken together, our results suggest that humans perceive hierarchically arranged motion scenes in close similarity to artificial observers of Bayesian structural inference. Our behavioral task should readily transfer to neuroscientific animal experiments to shed light on the neural representations of motion structure perception.

ADDITIONAL DETAILS

Motion structured stimuli. To probe the perception of latent motion structure in dynamic scenes, we employed the stochastic stimulus framework from [1] that composes hierarchical motion structures [2] from building blocks of varying strengths. Stimuli consisted of K=3 colored dots rotating on a circle (**Fig. 1A**) for 4s per trial, with dots following one of four motion structures S: (I) independent, (G) global, (C) clustered, or (H) hierarchically nested motion (**Fig. 1B**). Dot colors were randomized in each trial. Mathematically, dot velocities $\boldsymbol{v}_t$ followed a multivariate Ornstein-Uhlenbeck process, $d\boldsymbol{v}_t = -\boldsymbol{v}_t/\tau\, dt + \boldsymbol{L}\, d\boldsymbol{W}_t$, with time constant τ=1.5s and multivariate Wiener process $\boldsymbol{W}_t$ driving changes in velocity. The motion structure matrix $\boldsymbol{L}$ induces correlations in velocity across dots by distributing the components of $\boldsymbol{W}_t$ (e.g., global motion is achieved by a full column of identical values in $\boldsymbol{L}$, plus a smaller independent component for each dot). Every motion structure S in **Fig. 1B** can be represented by a matrix $\boldsymbol{L}$ (see [1] for details). Crucially, the resulting stationary distribution $p(\boldsymbol{v}_t, \boldsymbol{x}_t)$ of velocities and locations on the circle is known in closed form:

$$p(\boldsymbol{v}_t, \boldsymbol{x}_t) = \mathrm{Normal}(\boldsymbol{v}_t; \boldsymbol{0}, \boldsymbol{L}\boldsymbol{L}^T/2\tau) \cdot \prod_{k=1,2,3} \mathrm{Uniform}(x_{k,t}; 0, 2\pi)\,.$$

In the experiments, we chose the structures such that the marginal dot velocity distributions $p(v_{k,t})$, which are determined by the diagonal of $\boldsymbol{L}\boldsymbol{L}^T$, were identical across all tested motion structures S and dots k, thereby leaving correlations in velocity as the only information available to an observer about the motion structure S.

Experiment 1. 12 participants were briefly trained on the four motion structures in **Fig. 1B** and then performed 200 trials each in a 4-AFC task: after stimulus presentation, the participant selected the perceived structure and her/his confidence (high/low) in the decision. After the choice was made, the ground truth was revealed. The data of two representative participants is shown in **Fig. 1C** in the form of confusion matrices. The motion structures were designed such that humans often perceived the correct S, but also featured distinct error patterns when a stochastically generated scene was ambiguous (e.g., the I-C, H-G and H-C elements in **Fig. 1C**).

Bayesian observer model. For an observer, only the locations $\boldsymbol{x}_t$ are visible while velocities $\boldsymbol{v}_t$ and structure S need to be inferred. Like [1], we assume that observations are corrupted with Gaussian observation noise (std. σ) in every frame, modeling perceptual noise along the visual pathway. An ideal Bayesian observer would calculate the likelihood $p(\boldsymbol{x}_{1:T} \mid \mathrm{S})$ of the trial's trajectory $\boldsymbol{x}_{1:T}$ given motion structure S to make a decision. Since the Ornstein-Uhlenbeck process is linear, $p(\boldsymbol{x}_{1:T} \mid \mathrm{S})$ can be calculated using a Kalman filter with a motion structure-dependent transition model.

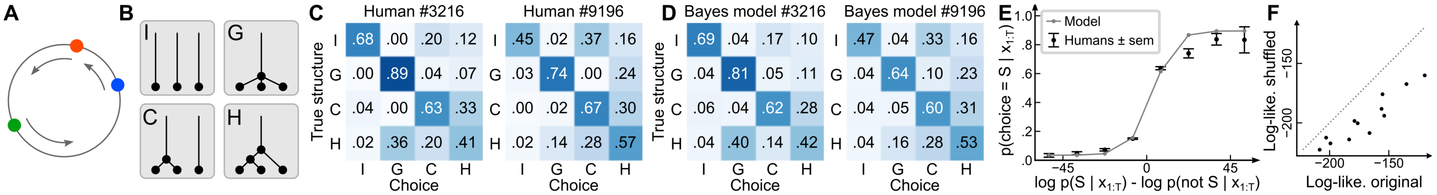
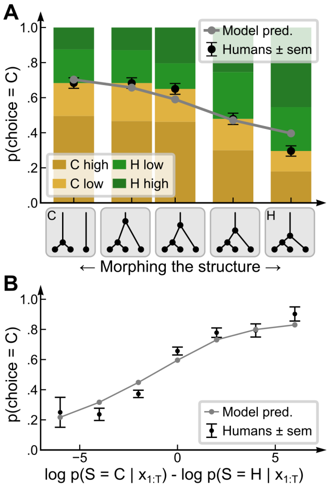


Figure 1: (A) Stimulus with global motion, (B) the four considered motion structures S, (C) confusion matrices of two participants, (D) Bayesian observer models' predictions, (E) psychometric curve for human choices (black) and model predictions (gray); (F) Shuffling trial-by-trial choices reduces quality of fitting (dots = participants).

Humans, however, are not ideal observers. Following [1,3,4], we therefore included a lapse probability π_{L} (random choice), an inverse temperature β (posterior-dependent stochastic choices) and biases b_{S} (prior over different structures) in our Bayesian observer model:

$$p(\,\text{choice} = \mathrm{S} \mid \boldsymbol{x}_{1:T}\,) = \pi_{\mathrm{L}}/4 + (1-\pi_{\mathrm{L}})\, e^{\beta\,\cdot\,(\log p(\,\boldsymbol{x}_{1:T}\mid \mathrm{S}\,)\,+\,b_{\mathrm{S}})} / \,(\text{Norm over all S}')\,,$$

where we correctly accounted for multiplicities due to color permutations by summing over equivalent structure sub-types (e.g., there are 3 sub-types for S=C, but only 1 sub-type for S=G; not included in the equation for clarity). For each participant, we then fitted four parameters $(\beta, b_{\mathrm{G}}, b_{\mathrm{C}}, b_{\mathrm{H}})$ via maximum likelihood of human choices under the observer model (b_{I}=0 by convention because the prior is normalized). The lapse probability π_{L}=0.14 and observation noise σ=1.1 were fitted as single values shared across all participants for better model identifiability. The resulting observer model with 4 per-participant and 2 shared parameters qualitatively predicted human choices (**Fig. 1D**; averaged over predictions obtained by repeatedly splitting the data into training and test sets, i.e., Monte-Carlo cross-validation). Further, we can calculate the psychometric response (**Fig. 1E**; summary for all participants) of human choices (black) and the fitted observer model (gray line) as a function of the log-posterior-odds of the ideal observer (x-axis), revealing the signature logistic shape of approximately correct Bayesian structural inference. Finally, we tested whether the observer model predicted human responses with single trial resolution by shuffling the human choices within each row of the confusion matrix, such that the confusion matrix itself remained unchanged, and refitted the observer model: maximum likelihood fits to the shuffled data achieved a significantly lower log-likelihood than fits to the unshuffled data (**Fig. 1F**), indicating that the observer model captures human choices with finer resolution than the "summary statistics" of the confusion matrices show in in **Figs. 1C & D**.



Experiment 2. To specifically study human motion structure perception of ambiguous scenes, we exploited the high confusability between the H and C structures (H-C element in Fig. 1C), by continuously varying the presented motion structure between the C and H prototypes (see x-axis of **Fig. 2A**) in a 2-AFC task. Experiment 2 directly followed experiment 1, with the same participants performing 100 trials each. But this time, they did not receive feedback and were not aware of the newly introduced intermediate motion structures. Human choices (bars in **Fig. 2A;** color hue = choice; lightness = confidence) gradually changed with the morphed stimulus. Strikingly, the observer model, which had been fitted to the first experiment, generalized to qualitatively predict human responses also to the often ambiguous, intermediate-structured stimuli of the second experiment (gray line). Replotting the data from Fig. 2A as a psychometric response function of the posterior-log-odds (**Fig. 2B**) confirms the expected logistic form. Furthermore, human confidence judgments c_{human} were correlated with the confidence proxy $c_{\text{ideal}} = |\, p(S = \mathrm{C} \mid \boldsymbol{x}_{1:T}) - 0.5\,|$ of the ideal observer (Pearson correlation $\rho(c_{\text{human}}, c_{\text{ideal}})$ = 0.11; p < 0.001).

Figure 2: (A) Human choices and confidence in the 2-AFC task (bars, black dots) alongside the predictions by the Bayesian observer model obtained from experiment 1 (gray line), (B) psychometric curve for human choices (black) and the model's predictions (gray line).

References. **[1]** Bill et al., *Hierarchical structure is employed by humans during visual motion perception*, bioRxiv (2019). **[2]** Gershman et al., *Discovering hierarchical motion structure*, Vision research (2016). **[3]** Drugowitsch et al., *Computational precision of mental inference as critical source of human choice suboptimality,* Neuron (2016). **[4]** Acerbi et al., *On the origins of suboptimality in human probabilistic inference*, PLoS CompBiol (2014).