Bistable perception emerges from loopy inference in strongly coupled probabilistic graphs

Alexandre Garcia-Duran (agarcia@crm.cat)
Centre de Recerca Matemàtica
Barcelona, Spain

Martijn Wokke
Universitat de Girona
Girona, Spain

Manuel Molano-Mazón
Universitat Politècnica de Catalunya
Barcelona, Spain

Alexandre Hyafil
Centre de Recerca Matemàtica
Barcelona, Spain

Abstract

During perception, the brain continuously processes sensory input, selecting among competing interpretations and assigning certainty -confidence- to each. Typically, confidence correlates with the strength of sensory evidence. In bistable perception, however, one interpretation is confidently perceived at a time, yet perception alternates despite no changes in the stimulus. We investigate which properties of visual stimuli drive this dissociation between evidence strength and confidence. We propose that bistability arises from approximate probabilistic inference over an internal representation of a stimulus with strongly coupled features. Using the Necker cube as an example, we model how perceived depth at each vertex is coupled with its neighbors, reflecting natural co-occurrence statistics. We analyze the dynamics of three inference algorithms. In all cases, strong feature coupling introduces loops in the internal representation that stabilize one percept, while internal noise drives perceptual switches. This creates a double-well potential, with perception fluctuating between high-confidence states. To test this, we designed a bistable stimulus in which feature coupling and sensory strength were independently manipulated. Our results show that stronger coupling leads to higher reported confidence, even when sensory evidence is weak. These findings suggest that bistable perception results from internal inference dynamics when stimulus features are tightly coupled.

Keywords: visual perception; bistable perception; perceptual inference; approximate inference; probabilistic graphical model

Approximate inference of bistable stimuli

Probabilistic graphical models (PGMs) offer a framework for understanding how inference runs between representations of low-level sensory features and large-scale objects. The PGM structure represents the probabilistic associations between features in the agent's internal model of the world (Gershman, Vul, & Tenenbaum, 2012).

Here, we used a binary Markov Random Field (MRF). In the Necker cube MRF (Fig. 1), the latent variables x_i (Fig. 1, circles) represent the depth of the corresponding vertices ($x_i = \pm 1$ front/back). The joint posterior distribution for each 3D configuration is $p(\mathbf{x}|s) \propto \exp(k(\mathbf{x}))$, with $\mathbf{x} = (x_1, ..., x_8)$, s the stimulus and $k(\mathbf{x}) = \sum_{(i,j)} \theta_{ij} x_i x_j + \sum_i B x_i$ the negative energy of each configuration. θ_{ij} represents the probabilistic coupling between features (Fig. 1, squares). Given that vertical and horizontal lines tend to join points at the same depth, we set $\theta_{ij} = J > 0$ when nodes i and j are connected horizontally/vertically, and $\theta_{ij} = -J$ if they are connected by a diagonal. B_i encodes the bias or sensory information. To compute the probability that a node is in the front or back, $p(x_i|s)$, and to reconstruct the 3D structure from the 2D image, we use approximate inference. This provides the approximate marginal $q(x_i) \approx p(x_i|s)$, representing confidence. Importantly, the same framework can be applied to model perception of other stimuli, simply adapting the MRF to reflect probabilistic associations between the stimulus features.

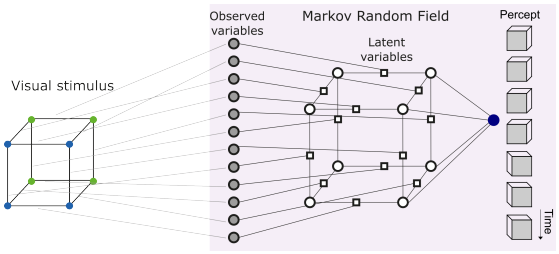

Figure 1: PGM of the Necker cube. The rightmost node represents a linear read-out of the marginal posteriors.

The coupling between beliefs leads to bistable perception

We investigated three common inference algorithms that have been hypothesized to be implemented by the sensory cortex (Haefner, Beck, Savin, Salmasi, & Pitkow, 2024): Gibbs Sampling (GS) (Gershman et al., 2012), Mean Field (MF) and Loopy Belief Propagation (LBP) —as well as its extension, Fractional Belief Propagation (Wiegerinck & Heskes, 2002)(FBP), designed to perform better in cyclic graphs. MF, LBP and FBP are embedded in a continuous dynamical system where each variable encodes one approximate marginal ($q_i(t) \approx q(x_i = 1)$). While MF assumes independence of the marginals over latent variables, LBP considers pairwise interactions. FBP weighs these interactions by a parameter α to prevent circular inference: $\alpha = 1$ corresponds to LBP, and $\alpha \to 0$ yields MF. In these algorithms the percept evolves within an energy landscape (Moreno-Bote, Rinzel, & Rubin, 2007) (Fig. 2a-b), with local minima as fixed points. Increasing the coupling (J) leads to a shift from monostable to bistable potential, whereas sensory evidence (B) modulates the asymmetry of the potential (Fig. 2a, middle). In the absence of sensory evidence, the critical coupling whereby bistability emerges is $J^* = 1/N$ for MF and $J^* = \log(N/(N-2\alpha))/2\alpha$ for FBP, where N is the number of node neighbours (N=3 for the Necker cube). Increasing α increases J^*, making inference more robust to loops, which reduces over-confidence.

On the other hand, GS updates one random binary variable x_i at each step n, where a change of value is more probable if it leads to a lower energy configuration: $p(x_i^n \neq x_i^{n-1} | \mathbf{x}^n, \mathbf{x}^{n-1}) = \sigma(k(\mathbf{x}^n) - k(\mathbf{x}^{n-1}))$, where $\sigma(x) = 1/(1 + \exp[-x])$. Under strong coupling, the system will spend more time in the least energetic states, corresponding to the two cube configurations, as the high energy barrier makes switches infrequent (Fig. 2c). The approximate posterior $q_i = q(x_i = 1)$ is taken as the average of x_i across T samples. We derive analytical approximations for the distribution of the posterior $p(q)$ (Fig. 2d). For short T, the system gets stuck in one interpretation, showing therefore a bimodal distribution. Only for very large T, the system explores the two interpretations equally, approximating the true posterior $p(q) = 0.5$.

If sensory evidence is added, then one interpretation will be more explored than the other.

We further derived analytically the stationary distribution for each algorithm, which allows us to compute model likelihood and to fit models to experimental data. All the algorithms also reproduce two hallmarks on bistable perception (not shown): Levelt's four propositions (Brascamp, Klink, & Levelt, 2015), which capture the relationship between the stimulus strength and the duration of dominance of a percept in bistable perception; and hysteresis (Hock, Kelso, & Schöner, 1993), a tendency of perception to be stuck temporarily even in the presence of an incongruent stimulus.

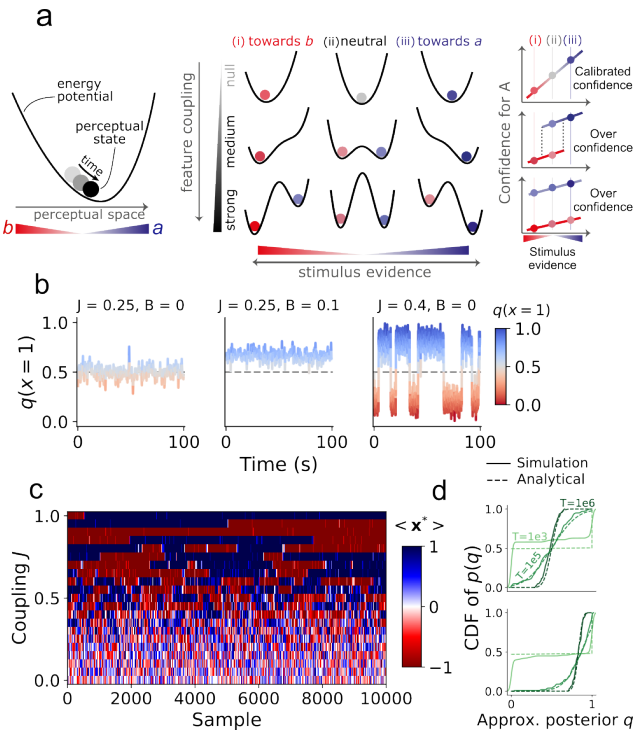

Figure 2: **a.** Left: perception evolves in an energy potential. Middle: This potential is modulated by coupling and stimulus evidence. Right: final percept. **b.** Example MF dynamics, for different values of J and B. **c.** Example GS dynamics, average of $\mathbf{x}^* = (x_1, ..., x_4, -x_5, ..., -x_8)$., for different values of J, $B = 0$. **d.** Cumulative distribution function (CDF) of the distribution of the posterior $p(q)$ for $B = 0$ (top) and $B > 0$ (bottom), $J = 1$.

Experimental validation

Human participants (N=30) were presented with random dots moving horizontally (green leftwards and red rightwards, or vice-versa), creating the perception of a rotating cylinder (Fig. 3a). They reported the color they perceive as being at the front as well as the clarity of their percept (i.e. perceptual confidence). Probabilistic coupling between the perception of depth of neighbouring dots is due to their similar velocity (structure-from-motion). Such coupling can be abolished by shuffling the velocity across dots. In the participants' reports, confidence increases with stimulus strength (relative size of red vs green dots; Fig. 3c, dots). Crucially, high coupling (zero shuffling) leads to overconfidence (irrespective of stimulus strength) (Fig. 3c bottom row, dots). In other words shuffling the velocity abolishes bistable perception by destroying coupling between features.

We then created a PGM for this particular stimulus (Fig. 3b), where latent variables also represent depth, and are coupled to their neighbors because of their similar velocity. After analyzing the inference algorithms in this framework, we fitted the MF model to the data (Fig. 3c, violins). The model could describe all the dependencies seen in the participants' reports. With this, we see that bistable perception emerges due to the probabilistic coupling between the features, which produce loopy inference.

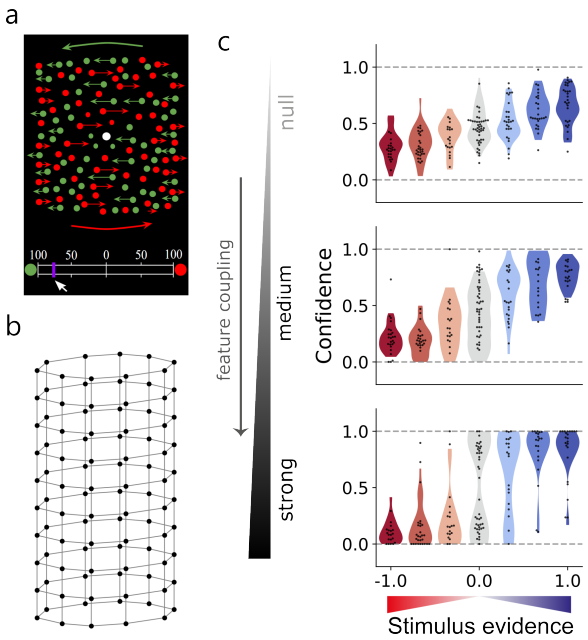

Figure 3: **a.** Experimental setup. **b.** PGM of the cylinder. **c.** Confidence towards green in front, of a single participant (dots), for different values of coupling and stimulus evidence. Violin plots represent the fitted model distributions.

References

Brascamp, J., Klink, P., & Levelt, W. (2015, April). The 'laws' of binocular rivalry: 50 years of levelt's propositions. *Vision Research*, *109*, 20–37. doi: 10.1016/j.visres.2015.02.019

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012, January). Multistability and perceptual inference. *Neural Computation*, *24*(1), 1–24. doi: 10.1162/neco$_{a0}$0226

Haefner, R. M., Beck, J., Savin, C., Salmasi, M., & Pitkow, X. (2024). How does the brain compute with probabilities? doi: 10.48550/ARXIV.2409.02709

Hock, H. S., Kelso, J. S., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 63–80. doi: 10.1037/0096-1523.19.1.63

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007, September). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, *98*(3), 1125–1139. doi: 10.1152/jn.00116.2007

Wiegerinck, W., & Heskes, T. (2002). Fractional belief propagation. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (Vol. 15). MIT Press.