Looking for hallmarks of generative models in the visual cortex

Gergő Orbán[1], **Pietro Berkes**[2], **Máté Lengyel**[3,4], **and József Fiser**[1]

[1]Brandeis University, [2]Gatsby Computational Neuroscience Unit, UCL,
[3]Collegium Budapest, [4]University of Cambridge

A recently emerging computational framework of the visual cortex assumes that it implements a generative model of (natural) visual input [1, 2]. According to this view, the visual cortex implicitly embodies a statistical model of how external causes (the latent variables of the model) combine to form the visual input (the observed variables of the model). Given a visual stimulus, the cortex inverts the model (according to Bayes' theorem), and thus infers which causes are likely to underlie it. Many psychophysical and physiological results are consistent with this hypothesis (see [3], for a review). However, testing this general idea directly is difficult, since it requires the correct specification not only of the generative model putatively implemented by the cortex, but also of its many implementational details.

An alternative approach is to look for fundamental hallmarks of generative models in the cortex that are not specific to any particular model, but are characteristic of probabilistic inference and generation and that require only minimal assumptions about the implementational details. We argue that one such hallmark of any generative model which adequately represents its input is a direct relationship between the prior distribution of latent variables, X, and their posterior distribution given some data present in the observed variables Y: $P(X) = \int P(X|Y)P(Y)dY$. Under the assumption that neural activity in the visual cortex represents samples from the distribution of latent variables [4], $P(X)$ and $P(X|Y)$ correspond to two different forms of V1 activity: that emerging in the absence of visual input (ie, spontaneous activity, SA), and that evoked by visual stimulus (EA), respectively. Thus, the above equation predicts that the statistics of SA must be identical to the statistics of EA (the latter integrated over a natural scene ensemble, $P(Y)$). Indeed, physiological recordings have shown that the statistics of EA movies in awake animals are remarkably similar to those recorded during SA [5].

Based on this framework, we analyze the activity of a hierarchical belief network, as a prototypical generative model, in order to identify other statistical hallmarks of generative models that can be found in the visual cortex. We examine the effects of probabilistic phenomena such as the relation between evoked and spontaneous activity, explaining away and contextual effects, and the effect of presenting noisy or ambiguous stimuli to the model. We discuss the kind of statistics that could be collected in in-vivo recordings in order to verify these effects, including measures based on data from a limited number of neurons.

Acknowledgments

This work was supported by a Sloan Swartz postdoctoral fellowship (GO), the Gatsby Charitable Foundation (PB), the EU Framework 6 (IST-FET 1940) (ML), and the NKTH (NAP2005/KCKHA005) (ML).

References

[1] Probabilistic Models of the Brain: Perception and Neural Function. RPN. Rao, BA Olshausen, and MS Lewicki, Eds. MIT Press, 2002.

[2] Vision as Bayesian inference: analysis by synthesis? A Yuille, D Kersten, *Tr Cogn Sci*, 10:301-8, 2006.

[3] Hierarchical Bayesian inference in the visual cortex. TS Lee, D Mumford, *J Opt Soc Am A Opt Image Sci Vis*, 20:1434-48, 2003.

[4] Interpreting neural response variablity as Monte Carlo sampling of the posterior. PO Hoyer and A Hyvarinen, In *Adv in Neural Inf Proc Systems*, 15: 277284. MIT Press, 2003.

[5] Small modulation of ongoing cortical dynamics by sensory input during natural vision. J Fiser, C Chiu, and M Weliky, *Nature* 431:573-8, 2004.