How to test Bayesian models using neural data

Gabor Lengyel (lengyel.gaabor@gmail.com)
Institute of Science and Technology Austria
Vienna, Austria
Department of Brain and Cognitive Sciences & Center for Visual Science, University of Rochester
Rochester, New York, United States of America

Sabyasachi Shivkumar (sabyashiv@gmail.com)
Zuckerman Institute, Columbia University
New York, New York, United States of America
Department of Brain and Cognitive Sciences & Center for Visual Science, University of Rochester
Rochester, New York, United States of America

Ralf M. Haefner (ralf.haefner@rochester.edu)
Department of Brain and Cognitive Sciences & Center for Visual Science, University of Rochester
Rochester, New York, United States of America

Abstract

The Bayesian Brain hypothesis suggests that the brain can be understood in terms of Bayesian computations. While many studies have provided perceptual and sensorimotor evidence for this hypothesis, the question of whether neural responses can also be understood in Bayesian terms remains open. Answering this question requires the specification of two main unknowns: (1) what is the generative model that relates the variables inferred by some population of neurons to the sensory observations, and (2) what is the "neural code", i.e. what is the relationship between posterior beliefs and neural responses? Much attention has been directed at answering the second of these questions while ignoring the first question, however without reaching consensus. At least in part this is because a given set of observed neural responses can imply different codes under different assumptions about the generative model. Here, we propose answering both questions in the opposite order. First, we present a method to test a given generative model using metamers – stimuli that give rise to the same posterior under this model – and confirming that they elicit the same neural responses. This approach can be interpreted as a special case of representational similarity analysis, and generalized accordingly. Second, we propose a "mixture method" that tests whether the relationship between posteriors for different stimuli matches the relationship for the measured neural responses to the same stimuli. If applied to the full response distribution, model and data are only expected to match for neural sampling codes. If applied to average neural responses, they are expected to match for any linear distributional code, including neural sampling and distributed distributional codes, but not probabilistic population codes. We illustrate our approach using simulations where the ground truth is known – both for a sparse coding model of V1, and a hierarchical motion model for area MT.

Keywords: Bayesian brain, perceptual inference, probabilistic population code, distributed distributional code, neural sampling code

Extended abstract

The empirical observation that sensory perception and sensorimotor behavior are often close to optimal (Knill & Pouget, 2004; Trommershäuser, Maloney, & Landy, 2008) has raised the question of whether the neural responses underlying those percepts and behavior can also be explained in Bayesian terms (Fiser, Berkes, Orbán, & Lengyel, 2010; Pouget, Beck, Ma, & Latham, 2013; Lange, Shivkumar, Chattoraj, & Haefner, 2023). If that is the case, then the relationship between stimulus (the brain's observation o) and neural responses, r, can be modeled as inference in a generative model $z \rightarrow o$ and a linking hypothesis about how the resulting posterior $p(z|o)$ is related to r ("neural code") (Fig. 1). No consensus exists about either ingredient, and most research to date has tried to test hypotheses about the representation (e.g. Neural Sampling Codes (NSC) (Fiser et al., 2010) vs Distributed Distributional codes (DDC) (Vértes & Sahani, 2018) vs Probabilistic Population Codes (PPC) (Ma, Beck, Latham, & Pouget, 2006)) making an assumption about the specific z represented by some neurons, or population of neurons: e.g. motion direction (Beck et al., 2008), intensity of image features (Hoyer & Hyvärinen, 2002; Orbán, Berkes, Fiser, & Lengyel, 2016; Haefner, Berkes, & Fiser, 2016), location of self (Ujfalussy & Orbán, 2022) etc. However, it has become clear that the conclusions drawn from this approach are sensitive to the assumed z, and that a different assumption about the nature of z (i.e. the generative model) might lead to a different conclusion (Shivkumar, Lange, Chattoraj, & Haefner, n.d.; Haefner, Beck, Savin, Salmasi, & Pitkow, 2024). The only existing approaches that avoided this problem tested the relationships between posteriors for different stimuli to the relationships between responses for the corresponding stimuli: either for natural stimuli (Berkes, Orbán, Lengyel, & Fiser, 2011) in a way that does not constrain the generative model, or for task-related stimuli in a way that constrains neither generative model nor the representation (Lange & Haefner, 2022). Our work generalizes and extends these prior approaches allowing for both the testing and comparison of candidate generative models, and for the testing and comparing of different hypotheses about the way probabilities are represented in neural responses.

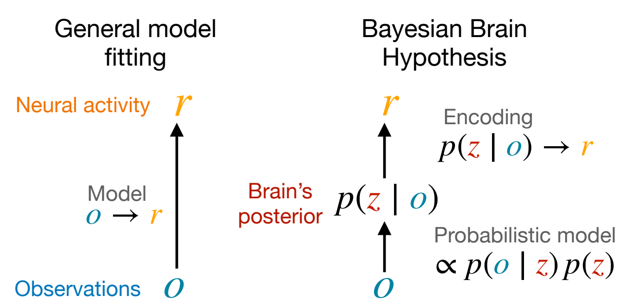

Figure 1: **Testing Bayesian models using neural data.** o, r, z, $p(\cdot)$ denote sensory observations, neural activity, brain's true latent variables, and probability distribution, respectively.

The *Metamers method* relies on the fact that for sufficiently complex generative models there will be many different stimuli/observations for which the (marginal) posterior over some latent variable (or subset of such variables) will be the same. For instance, in the case of a sparse coding model (Olshausen & Field, 1997), the posterior over a single latent feature will be the same for a wide range of stimuli with different combinations of orientation, spatial frequency, contrast, etc. (approximately all those for which the inner product of projective field and image is the same). Or for a motion model which hypothesizes that the brain computes relative motion, the posterior over relative velocity will be the same for many combinations of center and surround velocity (Fig. 2A). If the

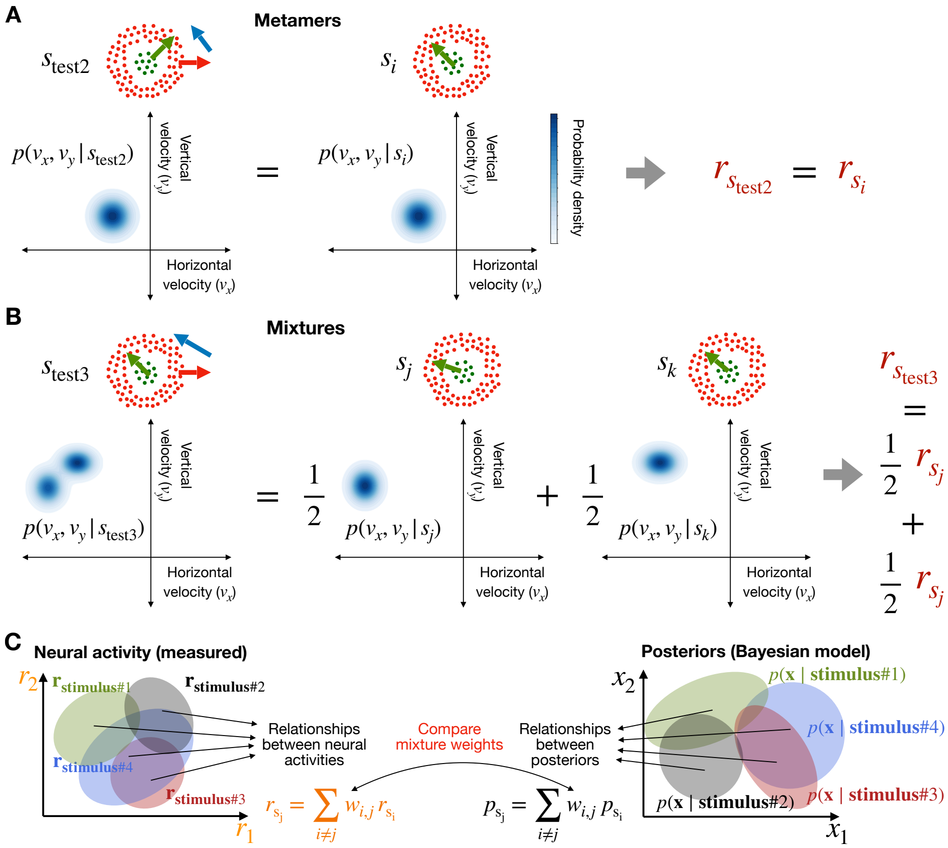

Figure 2: **Methods for generating neural predictions from Bayesian models. A-B:** The insets on top represent example center-surround motion stimuli. Colored arrows represent the motion of the center (green) and surround (red) dots, and the relative motion (center - surround, blue). v_x, v_y, $p(\cdot)$, $r_{\#}$, $S_{\text{test\#}}$, $S_{\#}$, denote horizontal and vertical velocities, probability distributions, neural activities, stimuli for which we generate neural predictions, and stimuli for which we measure the neural response that we use for generating the predictions. **C:** Mixture method, based on [anonymous author], if the encoding is linear distributional (NSC or DDC), relationships on the right should match those on the left.

Bayesian model accurately captures the probabilistic computations of a brain area, then the neural responses to these 'metamer' stimuli should be indistinguishable, regardless of how the posterior beliefs are encoded in the neural activity. For illustration, we applied this method to identify metamers, in a recently proposed Bayesian causal inference model of motion perception (Shivkumar, DeAngelis, & Haefner, 2023) (Fig. 2A). Our Metamer method can be thought of as a special case of representational similarity analysis (RSA), and hence generalized analogously: instead of testing whether neural responses are the same for stimuli designed to elicit the same posterior, a (dis)similarity matrix can be constructed between pairs of posteriors (e.g. based on their variational distance) corresponding to a large set of stimuli and comparing the matrix to the corresponding matrix based on neural responses as for RSA.

After confirming that a hypothesized generative model is compatible with the neural data, we propose testing the encoding of posterior beliefs with a method that relies on the distributional property of the respective probability encoding (Fig. 2B). NSCs are fully distributional in the sense that if the posterior to a specific stimulus j can be written as the mixture of the posteriors to a number of other stimuli $i \neq j$, then the same mixing weights w_{ij} should allow the neural response distribution to stimulus j be computed as a mixture of the response distributions to stimuli $i \neq j$ (Lengyel, Shivkumar, & Haefner, 2023). If the neural responses are based on a DDC or PPC, however, this equality will not hold. However, for DDCs, but not PPCs (which are nonlinear), the relationship will hold in expectation, i.e. the average response to stimuli j can be written as a linear combination of the average responses to stimuli $i \neq j$ (Fig. 2B,C). These properties allow for a discrimination between NSCs, DDCs, and PPCs using neural data.

Acknowledgments

This work was supported by the BRAIN Initiative grant from the National Institute of Neurological Disorders and Stroke (https://www.ninds.nih.gov/U19NS118246 to RMH) and by the NSF/CAREER grant (IIS-2143440).

References

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., . . . Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. , *60*(6), 1142–1152. doi: 10.1016/j.neuron.2008.09.021

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*(6013), 83-87. doi: 10.1126/science.1195870

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.*, *14*(3), 119–130.

Haefner, R. M., Beck, J., Savin, C., Salmasi, M., & Pitkow, X. (2024). *How does the brain compute with probabilities?* Retrieved from `https://arxiv.org/abs/2409.02709`

Haefner, R. M., Berkes, P., & Fiser, J. (2016). Perceptual Decision-Making as probabilistic inference by neural sampling. *Neuron*, *90*(3), 649–660.

Hoyer, P. O., & Hyvärinen, A. (2002). Interpreting neural response variability as monte carlo sampling of the posterior. In *Proceedings of the 15th international conference on neural information processing systems* (p. 293–300). Cambridge, MA, USA: MIT Press.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci*, *27*(12), 712–719.

Lange, R. D., & Haefner, R. M. (2022). Task-induced neural covariability as a signature of approximate Bayesian learning and inference. *PLoS Computational Biology*, *18*(3), e1009557. doi: 10.1371/journal.pcbi.1009557

Lange, R. D., Shivkumar, S., Chattoraj, A., & Haefner, R. M. (2023). Bayesian encoding and decoding as distinct perspectives on neural coding. *Nature Neuroscience*, *26*(12), 2063–2072. doi: 10.1038/s41593-023-01458-6

Lengyel, G., Shivkumar, S., & Haefner, R. M. (2023). A general method for testing bayesian models using neural data. In *Unireps: the first workshop on unifying representations in neural models.* Retrieved from `https://openreview.net/forum?id=oWJP0NhcY7`

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432-1438. doi: 10.1038/nn1790

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, *37*(23), 3311-3325. doi: https://doi.org/10.1016/S0042-6989(97)00169-7

Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and Sampling-Based probabilistic representations in the visual cortex. *Neuron*, *92*(2), 530–543.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci.*, *16*(9), 1170–1178.

Shivkumar, S., DeAngelis, G. C., & Haefner, R. M. (2023). Hierarchical motion perception as causal inference. *bioRxiv*. doi: 10.1101/2023.11.18.567582

Shivkumar, S., Lange, R. D., Chattoraj, A., & Haefner, R. M. (n.d.). A probabilistic population code based on neural samples. In *Advances of Neural Information Processig Systems 31* (pp. 7070–7079). Retrieved from `http://arxiv.org/abs/1811.09739`

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008, August). Decision making, movement planning and statistical decision theory. *Trends in cognitive sciences*, *12*(8), 291–7. doi: 10.1016/j.tics.2008.04.010

Ujfalussy, B. B., & Orbán, G. (2022, nov). Sampling motion trajectories during hippocampal theta sequences. *eLife*, *11*, e74058. doi: 10.7554/eLife.74058

Vértes, E., & Sahani, M. (2018). Flexible and accurate inference and learning for deep generative models. In *Advances in Neural Information Processing Systems* (pp. 4166–4175).