Neuron
Previews

CellPress

A probabilistic hammer for nailing complex neural data analyses

József Fiser[1,2,*] **and Ádám Koblinger**[1,2,*]
[1]Department of Cognitive Science, Central European University, Vienna, Austria
[2]Center for Cognitive Computation, Central European University, Budapest, Hungary
*Correspondence: fiserj@ceu.edu (J.F.), koblinger_adam@phd.ceu.edu (Á.K.)
https://doi.org/10.1016/j.neuron.2021.03.023

In this issue of *Neuron*, Młynarski et al. (2021) provide a maxent-based normative method for flexible neural data analysis by combining data-driven and theory-driven approaches. The next challenge is identifying the right frameworks to use this method at its best.

The art of interpreting any empirical data to gain a better insight about the underlying phenomenon has always been a balancing act between the Scylla of considering only the acquired information and the Charybdis of also using theoretically justified presumption about the processes that might have generated the data (Hastie et al., 2009). This dilemma is especially pertinent in investigating complex problems typical in biological and cognitive systems (Fiser et al., 2010). Bayesian data analysis has been accepted as the normative optimal method of maintaining this balance in the limit of a correctly defined model and infinite computational resources (Gelman et al., 2013). However, a major criticism of Bayesian methods is that the assumptions of the experimenter about the parameters and the correct model of the observed system are subjective (Bowers and Davis, 2012). This criticism motivated a prominent line of research on how to design priors that contain the least amount of presumptions about the estimated parameters (e.g., maxent priors) (Kass and Wasserman, 1996). Unfortunately, this approach reduces the unquestionable benefits of using pre-existing knowledge and assumptions about the parameters that are appropriate or could be essential for successfully exploring the data and the phenomenon behind it.

In this issue of *Neuron*, Młynarski et al. (2021) propose an approach to this conundrum and develop the corresponding technique, which rather than choosing between data-driven and initial-assumption-based approaches combines the benefits of the two. Młynarski and his colleagues start with the assumption that neural systems are highly optimized for particular tasks; thus, their internal model's parameters must be around an optimum. Therefore, instead of avoiding commitments about the model parameters, Młynarski et al. (2021) incorporate in their method very strong and normatively justified priors that reflect the presumed utility function of the underlying system. However, to accommodate chances that either the experimenter's model or the assumed utility might not perfectly capture reality, the authors introduce the idea that their chosen priors are not used *per se* as a determinant of a fixed set of parameters that reflect a pre-specified model. Instead, Młynarski et al. (2021) construct an optimization loop, in which they search probabilistically for the optimal interpretation based on a weighted combination of both data and the "appropriate" priors (Figure 1A). To achieve this weighting, they define a family of "optimization priors," priors that are not randomly defined over the parameter space but rather constrained to various degrees by the specific utility function that the system is assumed to be effectively optimized for. Among these priors, there is an extreme distribution that puts all the probability mass on the single parameter set that maximizes the expected utility of the model system and, correspondingly, restricts the parameter posterior to this single set (Figure 1B). This optimal parameter set is used as an anchor point for the remaining members of the family, which are defined around the anchor point by using the maximum entropy principle and various levels of preset average utility controlled by an "optimization parameter," a single scalar β. When $\beta = 0$, the relevance of parameters is uniformly distributed without any prior bias, but as $\beta \rightarrow \infty$, the parameters' distribution is gradually more biased toward the values defined by the anchor point and, hence, toward the model that is optimized for the presumed utility function.

Thus, this technique can build a continuous abstract trajectory in the model's parameter space with a "knob," a tunable parameter β that allows traveling along the trajectory between the best parameters reflecting only the available data and the best parameters of the selected model that could achieve the highest utility value with the given task (Figure 1B). This is an elegant solution to bridge data-driven and theory-driven approaches while considering various "intermediate" model settings between the two extremes in a principled manner. Being normative, the framework allows the use of the full repertoire of Bayesian computations for conducting inquiries relevant in the particular investigation. Młynarski et al. (2021) demonstrate this by considering four such inquiries on three different datasets: (1) performing a statistical test for optimality of a system based on given data and a presumed utility function, (2) inferring the system's degree of optimality, (3) disambiguating among various mappings between theoretical predictions and the parameters of the utility function, and (4) improving the inference in high-dimensional problems due to the structure of the optimization priors.

While the method of Młynarski et al. (2021) has the potential to help with analyzing complex data, it is worth putting

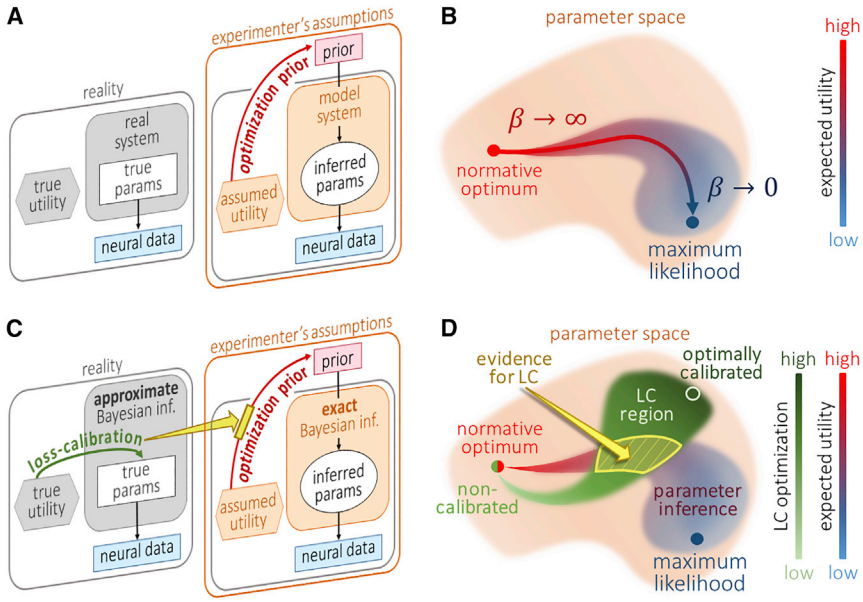

Figure 1. Optimization-prior-based parameter estimation and its integration with the idea of loss calibration

(A) The experimenter makes assumptions (orange frame) about the utility function, which determines the desired parametric state of the system, and about the structure of the real system that generates the experimenter's observations (neural data, blue rectangle). Optimization priors (red arrow) gradually introduce the experimenter's normative assumption that the system operates close to its optimum based on the assumed utility function.

(B) Applying the optimization priors effectively reduces the space of plausible model parameters during parameter inference. The optimality parameter, β, regulates the strength of optimization. Greater β values result in more concentrated posterior distributions (narrowing funnel) around higher utility parameter sets. Maximum *a posteriori* solutions for different β values (curved arrow) connects the normative optimum ($\beta = \infty$) with the maximum likelihood solution ($\beta = 0$).

(C) Our proposal of applying the optimization-prior method for testing the hypothesis of loss calibration (LC, green arrow). The costly computations with the approximate system (true system, gray rectangle) are replaced by simpler computations (yellow arrow), assuming that the system can perform exact Bayesian computations (model system, orange rectangle).

(D) LC designates a region in the approximate system's parameter space where the utility of the system is higher than the utility at the exact system's normative optimum (green funnel). Different points in this LC region correspond to different levels of optimization, including the two extremes where the system is not calibrated at all and where the system is optimally calibrated. This region can be mapped onto the parameter space of the exact system (orange cloud), potentially creating an overlap (yellow hatched region) with the optimization-prior-constrained region (red-blue funnel), providing evidence for LC in the approximate system. The presence of the overlap can be verified by evaluating the approximate model's utility along the optimization-prior-constrained region (red-blue funnel).

this framework into context by clarifying what it can provide as is and what aspects of the challenge still require additional development. Because of the structured nature of the optimization priors, the method can effectively and normatively exclude a huge fraction of the potential parameter space that would give rise to inadequate systems (models with specific parameters), and it can do this regardless of how well the particular systems are optimized for the given utility function. This is an important technical feat that opens the road to tackling a more complex class of data-analysis problems successfully based on limited amounts of data.

However, this technical tool relies heavily on the particular assumptions of the experimenter about both the right model of the investigated real system and the true utility the system aims to maximize (Figure 1A). As Młynarski et al. (2021) also point out in their discussion, the method does not help explicitly to tackle the two fundamental challenges of the analysis: identifying a proper utility function and defining a proper model (as opposed to just parameters) under which the system is close to optimal. Młynarski et al. offer an implicit way of using their methods for the first challenge (selecting the most appropriate utility function under which the model is the closest to being optimal) in simple cases using two standard Bayesian approaches: model selection for arbitrarily picked utility functions (e.g., sparseness versus slow feature change) and hierarchical parameter estimation for the continuous case.

While Młynarski et al. (2021) do not consider the second challenge, their probabilistic method can be applied in the complementary setup by systematically tuning the model structure to find the one that is the closest to be optimal given a particular utility function. In this alternative setup, optimality of two competing models could be investigated by model selection the same way as the optimality of the sparseness versus slow feature utility function was in the original setup. However, this extension highlights the method's limitation in feasibility for addressing the two fundamental challenges in realistic contexts: both the model system and the assumed utility function should be simple enough to make the model comparison or hierarchical parameter estimation tractable. While this limitation can be concealed in the case of finding the proper utility function by testing arbitrarily defined simple utility functions (such as sparseness), the complexity issue cannot be sidestepped so easily in the alternative case of searching for better models.

Here, we propose a systematic approach incorporating the method of Młynarski et al. (2021) to tackle rather than to avoid this problem of complexity in the case of searching for better models within the realm of probabilistic perceptual models. By relying on the utility function, an approximate probabilistic (Bayesian) inference can be performed so that it leads to the improvement of the expected utility of a resource-constrained system (Cobb et al., 2018). This is the idea behind loss-calibrated inference (LCI) (Figure 1C) (Lacoste-Julien et al., 2011). Testing whether processes in the brain follow LCI is not only important for finding the right models, but it is also a prime example of an investigation where the complexity of inferring the approximative models renders the simple

model selection techniques of Młynarski et al. (2021) unfeasible.

The extension we propose is loosely based on the idea of jumping between the true and the "proposal distribution" in sampling, where a more easily computable but different distribution is used for avoiding the expensive calculations with the true distribution. In our case, this amounts to inferring the parameters of the exact model with the method of Młynarski et al. (2021), which is cheaper than inferring the same parameters of the approximate model. The requirement in sampling that the proposal and the true distributions have to be proportional is equivalent in our case with the exact system and its approximate variants (sharing the same parameters) being similar. Using this setup, a comparison between the expected utilities of the approximate versions derived from the optimal exact model and from a model based on the authors' method can provide evidence for LCI. A higher expected utility at the latter would reflect a well-calibrated approximation on the real system, while a higher value at the optimal exact model would suggest that parameters identified by the method of Młynarski et al. (2021) represent just a generic deviation from the optimal anchor point.

The treatment of the above LCI problem is one example of how to capitalize on the appealing features of the new method of Młynarski et al. (2021) when analyzing real-world problems. The ultimate significance and impact of their framework hinges crucially on the success in integrating it with similar theory-driven analyses in the future.

REFERENCES

Bowers, J.S., and Davis, C.J. (2012). Bayesian just-so stories in psychology and neuroscience. Psychol. Bull. *138*, 389–414.

Cobb, A.D., Roberts, S.J., and Gal, Y. (2018). Loss-calibrated approximate inference in Bayesian neural networks. arXiv, 1805.03901, https://arxiv.org/abs/1805.03901.

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

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). Bayesian Data Analysis (CRC Press).

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer).

Kass, R.E., and Wasserman, L. (1996). The selection of prior distributions by formal rules. J. Am. Stat. Assoc. *91*, 1343–1370.

Lacoste-Julien, S., Huszár, F., and Ghahramani, Z. (2011). Approximate inference for the loss-calibrated Bayesian. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, *15*, G. Gordon, D. Dunson, and M. Dudík, eds., pp. 416–424.

Młynarski, W., Hledík, M., Sokolowski, T.R., and Tkačik, G. (2021). Statistical analysis and optimality of neural systems. Neuron *109*, S0896-6273(21)00042-8.