John Huguenard, Ph.D.

Senior Editor, *eLife*

Dear Dr. Huguenard,

We are pleased to resubmit our manuscript after several months of making extensive improvements to address the valid concerns of the reviewers. The reviewers indicated the need for an improved explanation of emergent property inference (EPI), comparisons to state-of-the-art alternative approaches, and greater depth and quality in the scientific analyses. In the revised manuscript, we explain that EPI is motivated by an incongruity between the practice of theoretical neuroscience and statistical inference techniques: the focus on emergent properties vs. data. Through this lens, we offer a more clearly written explanation of EPI, which is far more intuitive and less technical and saves such details for Methods.

The revised manuscript includes a thorough methodological comparison of EPI to modern likelihood-free inference techniques, including the method SNPE of the cosubmitted manuscript which has now been published by *eLife* (Goncalves et al. 2020). This comparison serves to prove two important points: a.) EPI can scale to higher dimensions than alternative techniques, and b.) EPI is the only inference technique that can precisely condition on emergent properties.

Finally, we have re-approached the manner in which models of primary visual cortex and superior colliculus are analyzed with EPI. This revised manuscript focuses on producing scientific insight from the *rich parametric structure* captured and quantified by deep probability distributions through EPI. Not only are these scientific insights obtained in a way that emphasizes the power of EPI, they are now much more appreciable with improved background context and supporting analyses.

Goncalves et al. 2020 was originally cosubmitted with this manuscript, and we have taken considerable effort and time to address important issues pointed out by the reviewers, while the cosubmission has been published. We are very supportive of this work, and recognize its publication date. While the two studies share the same high-level goals, they differ in the class of models and phenomena they target: Goncalves et al was motivated by the need to constrain, adjust and select mechanistic models of neural dynamics by experimental data, with a particular focus on models of ion-channels, cellular processes and biophysically detailed models of neural networks. On the other hand, this manuscript was motivated by the need to analyze circuit and systems-level models of neural dynamics which produce mathematical criteria or “emergent properties”, with a particular focus on cortical and recurrent neural network models. Futhermore, these methods differ greatly in their mathematical foundations, making the methodological comparison a highly valuable contribution to neuroscience and machine learning. We hope that if this manuscript is accepted, there is an opportunity for a highlight piece summarizing the contributions of the two works published at *eLife*.

Please do not hesitate to contact us if we can provide any further information.

Sincerely,

John P. Cunningham

on behalf of all authors.

submit two manuscripts for consideration in eLife. They are not under consideration elsewhere. We are submitting these manuscripts jointly as they address the same general question, and do so for distinct and complementary scientific purposes, and with different technical approaches (as we detail below). These manuscripts address a question of fundamental and timely importance to the neuroscience community: as theoretical models of computation become increasingly complex, how can we properly interrogate these models – their parameterizations, their degeneracies, scientific hypotheses generated by inspecting these models, and explanations of neural data generated by fitting these models to the wealth of data from modern recording techniques?

The current practice of this quantitative field often relies on outmoded and qualitative analyses, generating data from the model and making ad hoc statements about what parameters might have given rise to the emergent phenomenon under study. This state of affairs risks theoretical models - long the cornerstone of mathematical neuroscience, from the Hodgkin-Huxley model to chaotic recurrent neural networks – losing their position as the hallmark of quantitative, interpretable explanations of neural phenomena. Indeed, most recent quantitative assessments of neural data have been based on simple, purely phenomenological models (such as dimensionality reduction) which can be fit efficiently, but offer limited mechanistic insights.

We address this question by bringing deep learning to bear on the problem: in recent years, statistical machine learning has made tremendous progress in bringing rigorous inference to complex generative models. We show that we can modify and advance best-in-class machine learning technologies to these applications, and that it demystifies the operation of these black-box theoretical models. We present our methodologies and apply them alongside our experimental collaborators to a number of models of timely scientific importance.

During the preparation and preliminary presentation of our respective studies, we discovered that each of us was simultaneously and independently leading a project working towards the same high level goal of improving model-identification in neuroscience through deep inference tools. We are convinced that these methods will open up a new approach to modeling in neuroscience which is more systematic, powerful and efficient than the current approach to model-design in neuroscience, and which will be critically important to bridging the gap between theoretical and experimental approaches to studying neural function. We are similarly convinced that the magnitude of the challenge will require a multiplicity of tools aimed at different data and models, and that our studies have clear methodological and scientific differences, and make valuable and complementary contributions. Therefore, to ensure that their respective contributions can be reviewed and appreciated transparently, we decided to coordinate in efforts and submit our two studies in parallel. We note that we also have posted preprints to biorxiv simultaneously, which has generated significant excitement about the work.