We thank the reviewers for their thoughtful reviews and below we address their major concerns. We first emphasize that we will publish both the data and code if the paper is accepted—-this was an oversight by us for not making clear we would do so. As both Reviewers 1 and 2 point out, our current version of the state-state model uses task-specific ROIs identified from fMRI to inform the modeling of directed connectivity in EEG source space. Although the information from fMRI is incorporated only as a spatial prior, our simulation results in Section 3.1 demonstrate the importance of having accurate localization of ROIs (see blue bars for relative errors in Figure 2). It is clear that model performance decreases markedly even when a small number of sources are erroneously included in the ROIs. While it is possible to define ROIs from a separate fMRI experiment, one would expect increased variability in the shared neural activity between two non-simultaneously acquired modalities, leading to less accurate ROI boundaries and/or spurious ROI regions. This variability would be expected even from different recording sessions for the same subject. As a result, our model emphasizes the value of collecting simultaneous EEG-fMRI and using the two modalities to exploit their respective expressive power (spatial localization with fMRI and temporal dynamics in the EEG). Nevertheless, we agree with the reviewers that one might expect that the current model can be improved if fMRI information was also symmetrically integrated into the state-space framework.

In fact, in a previous version of this model, we did consider a generative process of fMRI which links to the EEG via the common latent dynamics $\mathbf{s}_t$. The fMRI BOLD signal was modeled as a linear convolution between the latent variable and canonical hemodynamic response functions (HRFs), which is expressed in linear matrix product form as:

$$\mathbf{z}_t[r] = [\mathbf{s}_t[r], \mathbf{s}_{t-1}[r], ..., \mathbf{s}_{t-L+1}[r]]', \;\; y_t^F[r] = \mathbf{w}'[r]\mathbf{H}\mathbf{z}_t[r] + \boldsymbol{\phi}_t^F[r] \tag{1}$$

where $\mathbf{z}_t[r]$ is a vector of L lagged values of the latent variable at the r^{th} ROI. $\mathbf{H}$ is a $3 \times L$ matrix that denotes a set of three hemodynamic basis functions, and $\mathbf{w}[r]$ is a 3×1 weight vector on the hemodynamic basis functions. Here we estimate a different $\mathbf{w}[r]$ for each ROI to account for the regional hemodynamic response variability. $\boldsymbol{\phi}_t^F[r]$ is i.i.d. Gaussian noise at the r^{th} ROI. Our simulation results for this model showed, however, that the addition of BOLD time series **does not** improve the estimation of latent dynamics compared to using EEG time series alone. This is not surprising since the temporal scale of the BOLD signal is much slower ($\sim$ 200 times slower) than that of the latent dynamics. It is the spatial specificity of the ROI localization offered by simultaneously recorded fMRI that contributes most to an accurate recovery of the latent dynamics. Another reason why we excluded the fMRI equation in the current paper is that we found using a linear convolution with canonical HRF functions did not predict the fMRI signals well. This is probably because the canonical HRF functions are estimated from a deconvolution between the task stimulus function (sparse events) and BOLD. We found that the BOLD signal is better predicted by the latent states convolved with an oscillatory-shaped HRF function, which substantially differs from the well-studied canonical form–this is an interesting finding but one that we still need to confirm and better understand. Since relatively little work has been done predicting BOLD from continuous EEG source activity, we believe that more investigation is needed before we report these findings.

Though our model shares some similarity with [7], it substantially generalizes this model and potentially has broader applications. First, our model is designed to explain the variance in continuously evolving EEG recordings as opposed to epoched EEG responses, as in [7]. This allows researchers to add multiple covariates (e.g., different experimental conditions) as external inputs into the dynamical system and investigate their individual influence. Second, our model can be easily applied to data not having a trial-based structure (e.g., resting-state data). We evaluated the model on task based EEG-fMRI data rather than resting state data because we originally wanted the model to be tested using a hypothesis-driven approach and our task EEG-fMRI data had been analyzed thoroughly by our group. In addition, for the resting-state case, there is no task activation so we potentially lose the ability to highlight the importance of the spatial specificity from fMRI. Nonetheless we agree with Reviewer 1 that more evaluation should be done using large-scale resting-state data. Also related to Reviewer 1's comments, it is certainly possible to have different numbers of ROIs for each subject. We chose the same number for each subject because these ROIs were obtained from a group-level fMRI activation and thus made group-inference on connectivity easier. Running our model with separate sets of ROIs per subject is possible, but a separate group inference procedure will be needed when subjects have different number of ROIs.

Another major point/question raised by the reviewers was the sensitivity of our results to our intialization procedure. Since variational inference algorithms are generally sensitive to the initialization, we chose a more informative initialization procedure which uses some fMRI-constrained techniques proposed in [12-14] to achieve a better initial solution for the EEG source localization. Using a completely random initialization is less likely to produce a good solution since the ELBO is expected to have many local optima. In our case, the final values substantially change from the initial values and move closer towards their true values; we will add a figure in the supplemental material showing this. In general, more detailed information about the initialization and algorithm runtime will be included in the supplement. **Minor concerns:** In the revision we will try our best to reorganize the material to include a more detailed description of the inference procedure in the main manuscript. We chose a diagonal $\boldsymbol{Q}_x$ as an approximation in the spirit of mean-field variational inference. It is not necessary but it simplifies the inference derivation.