We would like to thank all the reviewers for their thoughtful comments and their enthusiasm for our work. The reviewers' primary questions and concerns can be roughly condensed into four categories, which we address below.

Time resolution and latency between areas (R1, R4). These are very important considerations for maximizing the utility of our method for the neuroscience community. For simplicity, we only considered AR(1) models (i.e. $x_{t+1} = f(x_t)$), but the mp-srSLDS easily extends to AR(p) dynamics where many previous time bins are considered. By using AR(p) dynamics with small time bins, we can determine the latency of effects between populations. In fact, we already have an option for using AR(p) dynamics in our code, and will include an example in the Appendix. We would also like to mention that in our analyses of neural data, the interaction trends we observed were robust to varying time bin sizes (we tested 10, 25, and 50 ms in the motor cortex data). This may be because when using an AR(1) model, the dynamics may implicitly take more time scales into account by using additional latent dimensions to integrate over time.

Consequences of approximate posterior inference (R2, R4). How does the structured mean-field posterior approximation $p(z, x \mid y) \approx q(z)q(x)$ affect the inferred states and learned parameters? We found that Laplace EM with this mean-field approximation outperformed standard black box variational inference in the collapsed model obtained by summing over discrete states z, even though the collapsed model accounts for discrete and continuous state dependencies. These results are consistent with those of Zoltowski et al. [2020], where they found Laplace EM compared favorably to both BBVI and particle EM methods. We suspect these results reflect an inherent tradeoff between the fidelity of the posterior approximation and the difficulty of optimization, with simpler approximations (like Laplace EM with the structured mean-field approximation) sometimes leading to improved results [Turner and Sahani, 2011]. We will expand our discussion of these considerations in a camera-ready version.

Novel modeling contributions (R3). State space models, such as rSLDS models, form a strong foundation for many types of neural data analysis. However, the ability to easily interpret the interaction between multiple populations within rSLDS models was lacking. We explored three extensions to enhance interpretability, all of which are described within Section 3. Segmenting the continuous latent states for each population (which is equivalent to imposing hard constraints that the C matrix is block diagonal) simply and cleanly allows for per-population states and between-population interactions. On top of that, the "sticky" parameterization of discrete state transitions reveals which neural populations are responsible for staying in, or switching between, discrete states in the model. Finally, we developed further extensions that include more prior information on connectivity, which are discussed in both Section 3 and the Appendix. These contributions also pave the way for further investigation into how structural connectivity could be incorporated into prior distributions on multi-population interactions.

Findings from the analysis of neural data (R2, R3). *C. elegans* offers an illustrative demonstration of the mp-srSLDS as there are many possible definitions of 'population'. R3 questioned the value of 1 or 2-neuron populations, but for this organism, this is the level at which neuroscientists frequently study this circuit. The mp-srSLDS naturally handles this limiting case and reveals interactions between neuron classes (Section 5.3; Fig. 4), but it also admits other forms of population structure as well. For example, we explore interactions between ganglia in Appendix C.

Though one might expect strong feedforward influence of PMd on M1 during movement preparation, PMd has been shown to have a weaker influence on M1 during a preparatory phase [Kaufman et al., 2014] compared to during movement, in agreement with our results. Moreover, there are known feedback connections from M1 to PMd to produce the recurrent coupling we see during movement (and which R2 was curious about). More generally, in this dataset, our method allows seeing clear differences between inter-population dynamics during the movement and non-movement states (Section 5.2; Fig. 3), without precisely defining these states *a priori*.

There is a general challenge, shared among all descriptive statistical models, that modeling results do not provide causal insight on brain function. Our goal is that our method can 1) lead to a greater functional understanding, and 2) generate hypotheses that experimental neuroscientists can test with perturbation experiments.

Finally, for a camera-ready version, we will address all the minor concerns, including clarifying figures as suggested and adding missed citations. Thanks again for spending the time to provide valuable feedback on our work.

References

D. M. Zoltowski, J. W. Pillow, and S. W. Linderman. A general recurrent state space framework for modeling neural dynamics during decision-making. *Proceedings of the 37th International Conference on Machine Learning*, 2020.

R. E. Turner and M. Sahani. Two problems with variational expectation maximisation for time-series models. In *Bayesian Time series models*, chapter 5, pages 109–130. Cambridge University Press, 2011.

M. T. Kaufman, M. M. Churchland, S. I. Ryu, and K. V. Shenoy. Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440–448, 2014.