

We thank the reviewers for their analysis, appreciative words, and suggestions. Our reply below does not address the points that can be easily addressed by amending our manuscript.

Reviewer 1

1. ' $P(s)$ ': To be honest we are not sure what the Reviewer means with ' $P(s)$ -type constraints' (the Reviewer's comment is truncated). If it means the empirical frequency distribution of the sample activity, equivalent to n moment constraints, then it is one of the distributions calculated at step 2. of the protocol (lines 27–56), to be compared with the other distributions. We may add that using the empirical frequencies as constraints at the sample level is questionable: activities with zero frequencies lead to zeroes in the maximum-entropy distribution, and these are unreasonable because the sample size is much smaller than the state-space size. This indicates that maximum-entropy is used beyond its range of validity. The approach at the population level + marginalization cures this problem somewhat.

2. *Puzzling finds in the literature*: We apologize for having forgotten to discuss this; this oversight can be easily fixed.

Reviewer 2

1. *Comparison with other frameworks and long intro*: The comparison with other approaches and the exploration of more results are indeed important. There are two reasons why we gave less space to them:

a. We think that most literature hurries in deriving results and does not give enough space to the examination of the principles behind a method. This worries us. The misapplication of a method often cannot be detected by its results alone. This is especially true for maximum-entropy. We risk finding a plethora of results that may even look interesting but are meaningless, because the method was not meaningfully applied. In turn, this often happens because the problem we wanted to solve with that method is too vaguely defined. For this reason we wanted to give large space to defining the problem and to discussing how the method and its principles should be applied to it. Our wish is that our readers will give more thoughts to these matters.

b. We focus on the comparison between the 'naïve' maximum-entropy method, and the maximum-entropy method that accounts for the sampling context. The eight-page limit was not enough to make comparisons with more methods and discuss results. Also, a long discussion about further results would steal focus from the examination of the problem and the principles, which are our main concern and message.

2. *The effect of firing rates and correlations*: The empirical correlations determine (rather than affect) the result; see the protocol 1–3, lines 27–56. Firing rates surely affect the results, and this is an interesting question: most literature seem to seek a measure of 'cooperativity' that is invariant under firing-rate changes. We wonder whether maximum-entropy can provide this. Again, this is a question that should be addressed in defining the problem and the principles behind the method to solve it.

3. *Bayesian approaches and sampling*: The literature indeed offers fully Bayesian approaches, which we also prefer. But maximum-entropy methods still abound in the literature, and maybe they can be reasonably motivated. Our goal is to point out the subtlety in applying these methods. In the literature we have explored we've seen very little use of the basic sampling formulae presented in our paper. Honestly this surprised us, because they're surely essential even in a very basic analysis of the data.

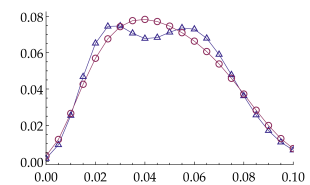
Reviewer 3

1. *Whether it matters in practice*: We agree with the Reviewer that the consequences of the differences between the two applications should be discussed further in the manuscript. In particular:

a. Minor differences may lead to considerably different interpretations of the underlying activity. For instance, the thinner right tail of the population-level distribution indicates that higher-order correlations observed in the data have larger statistical significance (this is the type of question Schneidman et al. were addressing in their research, which popularized maximum-entropy in neuroscience). From this point of view, a more entropic distribution is not necessarily more correct (after all, we could just use a uniform distribution otherwise).

b. Application of the method with higher-order constraints can show more interesting differences between sample- and population-level distributions. See e.g. the plot on the right, obtained with fourth-moment constraints: the bimodal distribution is the marginalized, population-level one.

c. Even if the differences were negligible, we believe that the results of the paper would matter: the point is that we could not have known about the presence or absence of differences, if we had not faced the whole problem and derived a formula showing that the difference were negligible.



2. *Unnecessarily principled*: For our motivation in emphasizing principles, we kindly ask the Reviewer to read our reply 1. to Reviewer 2 above. Probably the writing could have improved on our part to make the paper flow more smoothly, without sacrificing the more philosophical discussion?