We thank the reviewers for their analysis, appreciative words, and suggestions.

**Reviewer 1**

1. It would be interesting to use the frequency distribution of the sample activity, measured in the recording, as a constraint, equivalent to $n$ moment constraints. In fact, using it at the population level is less questionable than using at the sample level (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). Its use, however, addresses a different question than the one discussed in our introduction, namely to define and measure 'cooperativity' by comparing maximum-entropy distributions constructed from different numbers of moments.

2. Unfortunately the reviewer forgot the bibliographic reference to Tkacik et al 2012. We'd like to read that paper.

3. It's true, we forgot to discuss how the population-level application solves the sample-size-dependence of maximum-entropy results pointed out in previous literature. This should be addressed in an amended version of the paper.

**Reviewer 2**

1. We agree that the comparison with other approaches is important. But we believe that it's also important to carefully examine (1) the question we're asking, and (2) which first principles, if any, can be used to translate it into a mathematical problem. A lot of literature develops very refined – though often ad hoc – mathematical techniques; and yet it leaves us unsatisfied, with the lingering question 'why were we doing this?'. This is the reason for our long introductory discussion about the problem and the first principles that can be used to address it.

2. Regarding the effect of correlations on the result: the original question is how 'cooperativity' can be related to the measured correlations, or defined in terms of them. The method is devised to give one answer to this question; see the 'protocol' 1–3 in the Introduction. The crucial point is that there is no underlying distribution; the point of the method is to construct it for us.

3. We agree that there are many approaches in the literature. 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. We also agree with Reviewer 3 that the main formulae of the paper, (17) and (19), probably have little experimental use today. But the point they make is important:

   - the Reviewer says 'the differences between the sample level and population level distribution seem minor' – but we could not have known this, if we hadn't faced the whole problem and derived a formula showing that the difference is minor;
   - the sample-level distribution is actually *less* entropic. By lifting the constraints to the population level, the maximum-entropy principle allows us to explore even *more entropic* sample distributions than those allowed by a direct application at the sample level. Thus this approach agrees with the 'maximum entropy' spirit. – These points should be mentioned, of course, in an amended version of the paper.
   -

2. We agree with Reviewer 3 that the paper emphasizes some philosophical aspects; intentionally so. Sadly we can't make all kinds of readers happy; a choice of audience is necessary. As readers ourselves we appreciate when a paper begins by asking: 'What is the question? is it possible to translate it into a mathematical problem? which principles can we use to make such translation?'. There are papers that develop very refined mathematical techniques but leave us unsatisfied, with the lingering question 'why are we doing this?'. This is the reason why we try to emphasize these kinds of questions. But we are sure that part of the NIPS audience will appreciate this emphasis.