Key insight is that statistical dynamics steps give us a target that we can use. This causes bootstrap.

Noe etc use cases with some training data. We try to focus on cases where we only have log prob function. (Note it is possible to incorporate their training methods in, but useful to focus on har far we can get without this data, as it is very expensive to get).