table belong to the same cluster. Since the number of occupied tables is random, this provides a flexible model in which the number of clusters is determined by the data.

The customers of a CRP are exchangeable—under any permutation of their ordering, the probability of a particular configuration is the same—and this property is essential to connect the CRP mixture to the DP mixture. The reason is as follows. The Dirichlet process is a distribution over distributions, and the DP mixture assumes that the random parameters governing the observations are drawn from a distribution drawn from a Dirichlet process. The observations are conditionally independent given the random distribution, and thus they must be marginally exchangeable.[1] If the CRP mixture did not yield an exchangeable distribution, it could not be equivalent to a DP mixture.

Exchangeability is a reasonable assumption in some clustering applications, but in many it is not. Consider data ordered in time, such as a time-stamped collection of news articles. In this setting, each article should tend to cluster with other articles that are nearby in time. Or, consider spatial data, such as pixels in an image or measurements at geographic locations. Here again, each datum should tend to cluster with other data that are nearby in space. While the traditional CRP mixture provides a flexible prior over partitions of the data, it cannot accommodate such non-exchangeability.

In this paper, we develop the *distance dependent Chinese restaurant process*, a new CRP in which the random seating assignment of the customers depends on the distances between them.[2] These distances can be based on time, space, or other characteristics. Distance dependent CRPs can recover a number of existing dependent distributions (Ahmed and Xing, 2008; Zhu et al., 2005). They can also be arranged to recover the traditional CRP distribution. The distance dependent CRP expands the palette of infinite clustering models, allowing for many useful non-exchangeable distributions as priors on partitions.[3]

The key to the distance dependent CRP is that it represents the partition with *customer assignments*, rather than table assignments. While the traditional CRP connects customers to tables, the distance dependent CRP connects customers to other customers. The partition of the data, i.e., the table assignment representation, arises from these customer connections. When used in a Bayesian model, the customer assignment representation allows for a straightforward Gibbs sampling algorithm for approximate posterior inference (see Section 3). This provides a new tool for flexible clustering of non-exchangeable data, such as time-series or spatial data, as well as a new algorithm for inference with traditional CRP mixtures.

Related work. Several other non-exchangeable priors on partitions have appeared in recent research literature. Some can be formulated as distance dependent CRPs, while others represent a different class of models. The most similar to the distance dependent CRP is the probability distribution on partitions presented in Dahl (2008). Like the distance dependent CRP, this distribution may be

1. That these parameters will exhibit a clustering structure is due to the discreteness of distributions drawn from a Dirichlet process (Ferguson, 1973; Antoniak, 1974; Blackwell, 1973).
2. This is an expanded version of our shorter conference paper on this subject (Blei and Frazier, 2010). This version contains new perspectives on inference and new results.
3. We avoid calling these clustering models "Bayesian nonparametric" (BNP) because they cannot necessarily be cast as a mixture model originating from a random measure, such as the DP mixture model. The DP mixture is BNP because it includes a prior over the infinite space of probability densities, and the CRP mixture is only BNP in its connection to the DP mixture. That said, most applications of this machinery are based around letting the data determine their number of clusters. The fact that it actually places a distribution on the infinite-dimensional space of probability measures is usually not exploited.