restaurant franchise (CRF) which we refer to as a *CRF with loyal customers*. Here, each restaurant in the franchise has a specialty dish with the same index as that of the restaurant. Although this dish is served elsewhere, it is more popular in the dish's namesake restaurant. Recall that while customers in the CRF of the HDP are pre-partitioned into restaurants based on the fixed group assignments, in the HDP-HMM the value of the state z_t determines the group assignment (and thus restaurant) of customer y_{t+1}. The increased popularity of the house specialty dish (determined by the sticky parameter κ) implies that children are more likely to eat in the same restaurant as their parent ($z_t = z_{t-1} = j$) and, in turn, more likely to eat the restaurant's specialty dish ($z_{t+1} = j$). This develops family loyalty to a given restaurant in the franchise. However, if the parent chooses a dish other than the house specialty ($z_t = k$, $k \neq j$), the child will then go to the restaurant where this dish is the specialty and will in turn be more likely to eat this dish, too. One might say that for the sticky HDP-HMM, children have similar taste buds to their parents and will always go to the restaurant that prepares their parent's dish best. Often, this keeps many generations eating in the same restaurant.

Throughout the remainder of the paper, we use the following notational conventions. Given a random sequence $\{x_1, x_2, \ldots, x_T\}$, we use the shorthand $x_{1:t}$ to denote the sequence $\{x_1, x_2, \ldots, x_t\}$ and $x_{\backslash t}$ to denote the set $\{x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T\}$. Also, for random variables with double subindices, such as $x_{a_1 a_2}$, we will use $\mathbf{x}$ to denote the entire set of such random variables, $\{x_{a_1 a_2}, \forall a_1, \forall a_2\}$, and the shorthand notation $x_{a_1\cdot} = \sum_{a_2} x_{a_1 a_2}$, $x_{\cdot a_2} = \sum_{a_1} x_{a_1 a_2}$ and $x_{\cdot\cdot} = \sum_{a_1} \sum_{a_2} x_{a_1 a_2}$.

5.1. *Sampling via direct assignments.* In this section we present an inference algorithm for the sticky HDP-HMM of Section 5 and Figure 5(a) that is a modified version of the direct assignment Rao-Blackwellized Gibbs sampler of Teh et al. (2006). This sampler circumvents the complicated bookkeeping of the CRF by sampling indicator random variables directly. The resulting sticky HDP-HMM direct assignment Gibbs sampler is outlined in Algorithm 1 of the Supplementary Material [Fox et al. (2010)], which also contains the full derivations of this sampler.

The basic idea is that we marginalize over the infinite set of state-specific transition distributions π_k and parameters θ_k, and sequentially sample the state z_t given all other state assignments $z_{\backslash t}$, the observations $y_{1:T}$, and the global transition distribution β. A variant of the Chinese restaurant process gives us the prior probability of an assignment of z_t to a value k based on how many times we have seen other transitions from the previous state value z_{t-1} to k and k to the next state value z_{t+1}. As derived in the Supplementary Material [Fox et al. (2010)], this conditional distribution is dependent upon whether either or both of the transitions z_{t-1} to k and k to z_{t+1} correspond to a self-transition, most strongly when $\kappa > 0$. The prior probability of an assignment of z_t to state k is then weighted by the likelihood of the observation y_t given all other observations assigned to state k.