APPENDIX A
CULINARY METAPHOR

Following a similar procedure as in [1], we can derive the expression for \( \lim_{M \to \infty} p(\{S\}|Q, \alpha, \beta_0, \beta, \gamma) \) given in the Section 2.2 of the main text from a culinary metaphor analogous to the IBP. In this process, \( T \) customers enter sequentially a restaurant with an infinitely long buffet of dishes. The first customer starts at the left of the buffet and takes a serving from each dish, taking (possibly different) quantities for each one and stopping after a Poisson \( \{Q\} \) number of dishes. The number of units \( q \) for each one and stopping after a Poisson \( \{Q\} \) number of dishes. The number of units \( q \in \{1, \ldots, Q - 1\} \) she takes is independently sampled for each dish from a uniform distribution.

The \( t \)-th customer enters the restaurant and starts at the left of the buffet. At dish \( m \), she looks at the customer in front of her to see how many units she has taken from that dish and proceeds as follows:

- If the \( (t - 1) \)-th customer did not take the \( m \)-th dish, she serves herself that dish with probability

  \[
  \frac{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}}{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}} \quad (i = 1, \ldots, Q - 1),
  \]

  or she does not take that dish with probability

  \[
  \frac{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}}{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}} \quad (i = 1, \ldots, Q - 1).
  \]

- If the \( (t - 1) \)-th customer took \( q \) units from the \( m \)-th dish, the \( t \)-th customer either serves herself \( i \) units with probability

  \[
  \frac{\beta_0 (Q - 1) + \sum_{j=0}^{Q-1} n_{qi}^{m, j}}{\beta_0 (Q - 1) + \sum_{j=0}^{Q-1} n_{qi}^{m, j}} \quad (i = 1, \ldots, Q - 1),
  \]

  or she does not take that dish with probability

  \[
  \frac{\beta_0 (Q - 1) + \sum_{j=0}^{Q-1} n_{qi}^{m, j}}{\beta_0 (Q - 1) + \sum_{j=0}^{Q-1} n_{qi}^{m, j}} \quad (i = 1, \ldots, Q - 1),
  \]

  \[
  \frac{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}}{\beta_0 + \sum_{j=0}^{Q-1} n_{qi}^{m, j}} \quad (i = 1, \ldots, Q - 1).
  \]

The \( t \)-th customer then moves on to the next dish and repeats the above procedure. After having passed all dishes people have previously served themselves from, she takes independent quantities \( q \sim \) Uniform \( \left( \frac{1}{Q - 1}, \ldots, \frac{1}{Q - 1} \right) \) from a Poisson \( \{Q\} \) number of new dishes.

If we fill in the entries of the \( T \times M \) matrix \( S \) with the number of units that every customer took from every dish, and we denote with \( M_1^{(t)} \) the number of new dishes tried by the \( t \)-th customer, the probability of any particular matrix \( S \) being produced by this process is given in Eq. 1.

It is straightforward to check that there are

\[
\frac{(Q - 1)!}{(Q - N_{Q})!} \prod_{t=1}^{T} M_1^{(t)}! / \prod_{h=1}^{Q - 1} M_1^{(t)}! \]

matrices in the same equivalence class as \( S \), and therefore we can recover \( \lim_{M \to \infty} p(\{S\}|Q, \alpha, \beta_0, \beta, \gamma) \) by summing over all possible matrices lying in the set \( \{S\} \) generated by this process.

APPENDIX B
ASSIGNMENT PROBABILITIES FOR THE GIBBS SAMPLER

We now derive the probability \( p(s_{tm} = k|S_{-tm}) \), needed in Section 4.1 of the main text. This expression can be expressed, up to a proportionality constant, as shown in Eq. 2. Let \( n_{qi}^{m, tm} \) be the number of transitions from state \( q \) to state \( i \) in chain \( m \), excluding the transitions from state \( s_{(t-1)m} \) to \( s_{tm} \) and from state \( s_{tm} \) to \( s_{(t+1)m} \). Similarly, let \( n_{q}^{m, tm} \) be the total number of transitions from state \( q \) in chain \( m \) without taking into account state \( s_{tm} \), namely, \( n_{q}^{m, tm} = \sum_{i=0}^{Q-1} n_{qi}^{m, tm} \). The expression in (2) takes different forms depending on the values of \( j = s_{(t-1)m} \) and \( \ell = s_{(t+1)m} \), yielding Eq. 3 for \( j = 0 \) and Eq. 4 for \( j \neq 0 \).
APPENDIX C
UPDATE EQUATIONS FOR THE VARIATIONAL ALGORITHM

The variational inference algorithm involves optimizing the variational parameters of $q(\Psi)$ to minimize the Kullback-Leibler divergence of $q$ in the variational parameters of $p_M(\Psi | \mathcal{X}, \mathcal{H})$ from $q(\Psi)$, i.e., $D_{KL}(q || p_M)$. This optimization can be performed by iteratively applying the following fixed-point set of equations:

$$P^{m}_{jk} = \begin{cases} 
\exp \left\{ \psi(t^m_{jk}) - \psi \left( \sum_{i=0}^{Q-1} t^m_{ji} \right) \right\}, & \text{if } j \neq 0 \\
\exp \left\{ \psi(\nu^m_1) - \psi(\nu^m_1 + \nu^m_2) \right\}, & \text{if } j = 0, k = 0 \\
\exp \left\{ \psi(\nu^m_2) - \psi(\nu^m_1 + \nu^m_2) + \psi(\epsilon^m_k) \right\} - \psi \left( \sum_{i=1}^{Q-1} \epsilon^m_i \right), & \text{if } j = 0, k \neq 0 
\end{cases}$$

(5)

$$b^{m}_{kt} = \exp \left\{ -\frac{1}{2\sigma^2} (L_k)_m (L_k)^\top \right\} + \frac{1}{\sigma^2} (L_k)_m \left( x_t - \sum_{\ell \neq m} \sum_{i=1}^{Q-1} (L_i)_\ell q(s_{t\ell} = i) \right)^\top$$

(6)

$$\tau^m_{jk} = \delta_{k0} \lambda_0 + (1 - \delta_{k0}) \beta + \sum_{t=1}^{T} q(s_{(t-1)m} = j, s_{tm} = k)$$

(7)

$$\nu^m_1 = 1 + \sum_{t=1}^{T} q(s_{(t-1)m} = 0, s_{tm} = 0)$$

(8)

$$\nu^m_2 = \frac{\alpha}{M} + \sum_{t=1}^{T} q(s_{(t-1)m} = 0, s_{tm} > 0)$$

(9)

$$\epsilon^m_k = \gamma + \sum_{t=1}^{T} q(s_{(t-1)m} = 0, s_{tm} = k)$$

(10)

$$\Omega_k = \left( \frac{1}{\sigma^2_0} + \frac{M}{\sigma^2_\phi} \right)^{-1} I_D$$

(11)

$$\omega_k = \left( \frac{1}{\sigma^2_\phi} I_M L_k + \frac{1}{\sigma^2_0} \mu_0 \right) \Omega_k$$

(12)

$$\Lambda_k = \left( \frac{1}{\sigma^2_\phi} I_M + \frac{1}{\sigma^2_x} C_k \right)^{-1}$$

(13)

$$L_k = A_k \left( \frac{1}{\sigma^2_\phi} 1_M \omega_k + \frac{1}{\sigma^2_x} Q_k^\top \left( X - \sum_{j \neq k} Q_j L_j \right) \right)$$

(14)

where $(L_k)_m$ denotes the $m$-th row of matrix $L_k$, $\psi(\cdot)$ stands for the digamma function [2, p. 258–259], $\delta_{i'j'}$ denotes the Kronecker delta function (which takes value one if $i = i'$ and zero otherwise), and the elements of the $T \times M$ matrices $Q_k$ and $M \times M$ matrices $C_k$ are, respectively, given by

$$(Q_k)_{tm} = q(s_{tm} = k)$$

(15)

and

$$(C_k)_{mm'} = \begin{cases} 
\sum_{t=1}^{T} q(s_{tm} = k)q(s_{tm'} = k), & \text{if } m \neq m' \\
\sum_{t=1}^{T} q(s_{tm} = k), & \text{if } m = m'. 
\end{cases}$$

(16)

The probabilities $q(s_{tm})$ and $q(s_{tm}, s_{(t-1)m})$ can be obtained through a standard forward-backward algorithm for HMMs within each chain, in which the variational parameters $P^{m}_{jk}$ and $b^{m}_{kt}$ play respectively the role of the transition probabilities and the observation probability associated with state variable $s_{tm}$ taking value $k$ in the Markov chain $m$ [3].

REFERENCES


\[ p(S|Q, \alpha, \beta_0, \gamma) = \frac{\alpha^{M_+}}{M_{11}(t)!} \prod_{t=1}^{T} e^{-\alpha H_T} \times \prod_{m=1}^{M_+} \left[ \frac{\Gamma(n_{0i}^{\text{tm}} + 1) \Gamma \left( \sum_{i=1}^{Q-1} n_{0i}^{\text{tm}} \right) \Gamma \left( \left( Q-1 \right) \gamma \right) \prod_{i=1}^{Q-1} \Gamma(n_{0i}^{\text{tn}} + \gamma) \Gamma \left( \sum_{i=1}^{Q-1} \left( n_{0i}^{\text{tm}} + \gamma \right) \right) \Gamma(\gamma) \prod_{q=1}^{Q-1} \left( \Gamma(\beta_0 + (Q-1)\beta) \prod_{i=1}^{Q-1} (\Gamma(\beta_0^{\text{tm}} + \beta_0) \prod_{i=1}^{Q-1} (\Gamma(n_{0i}^{\text{tn}} + \beta)) \right) \right] \] 

(1)

\[ p(s_{tm} = k|S_{-tm}) \propto \begin{cases} \int_{a_k^m} p(s_{tm} = k|a_k^m)p(a_k^m|\{s_{\tau m}|s_{(\tau-1)m} = j, \tau \neq t, t + 1\}) da_k^m \times \\
\times \int_{a_k^m} p(s_{(t+1)m} = \ell|a_k^m)p(a_k^m|\{s_{\tau m}|s_{(\tau-1)m} = k, \tau \neq t, t + 1\}) da_k^m, & \text{if } k \neq j \\
\int_{a_k^m} p(s_{(t+1)m} = \ell, s_{tm} = k|a_k^m)p(a_k^m|\{s_{\tau m}|s_{(\tau-1)m} = k, \tau \neq t, t + 1\}) da_k^m, & \text{if } k = j. \end{cases} \]

(2)

a) If \( j = 0 \):

\[ p(s_{tm} = k|S_{-tm}) \propto \begin{cases} \frac{n_{00}^{\text{tm}} + 1}{(n_{0*}^{\text{tm}} + 1)(n_{0*}^{\text{tm}} + 2)} \left( \delta_{t0}(n_{00}^{\text{tm}} + 2) + (1 - \delta_{t0}) \frac{(\gamma + n_{0t}^{\text{tm}}) \sum_{i=1}^{Q-1} n_{0i}^{\text{tm}}}{\sum_{i=1}^{Q-1} (\gamma + n_{0i}^{\text{tm}})} \right), & \text{if } k = 0 \\
(\delta_{t0}\beta_0 + (1 - \delta_{t0})\beta + n_{k\ell}^{\text{tm}})(\gamma + n_{0\ell}^{\text{tm}}) \left( \sum_{i=1}^{Q-1} n_{0i}^{\text{tm}} \right), & \text{if } k = 1, \ldots, Q-1. \end{cases} \]

(3)

b) If \( j \neq 0 \):

\[ p(s_{tm} = k|S_{-tm}) \propto \begin{cases} \frac{(\beta_0 + n_{j0}^{\text{tm}})}{(n_{0*}^{\text{tm}} + 1)(\beta_0 + (Q-1)\beta + n_{j*}^{\text{tm}})} \left( \delta_{t0}(n_{00}^{\text{tm}} + 1) + (1 - \delta_{t0}) \frac{(\gamma + n_{0t}^{\text{tm}}) \sum_{i=1}^{Q-1} n_{0i}^{\text{tm}}}{\sum_{i=1}^{Q-1} (\gamma + n_{0i}^{\text{tm}})} \right), & \text{if } k = 0 \\
(\delta_{t0}\beta_0 + (1 - \delta_{t0})\beta + n_{k\ell}^{\text{tm}} + \delta_{k\ell}\delta_{jk})(\beta + n_{j\ell}^{\text{tm}}) \left( \sum_{i=1}^{Q-1} \gamma + n_{0i}^{\text{tm}} \right), & \text{if } k = 1, \ldots, Q-1. \end{cases} \]

(4)