A Bayesian Analysis of Dynamics in Free Recall

Richard Socher[1], Samuel Gershman[2], Adler Perotte[2], Per Sederberg[2], David Blei[2] & Kenneth Norman[2]
[1]Stanford University, [2] Princeton University

1 Overview

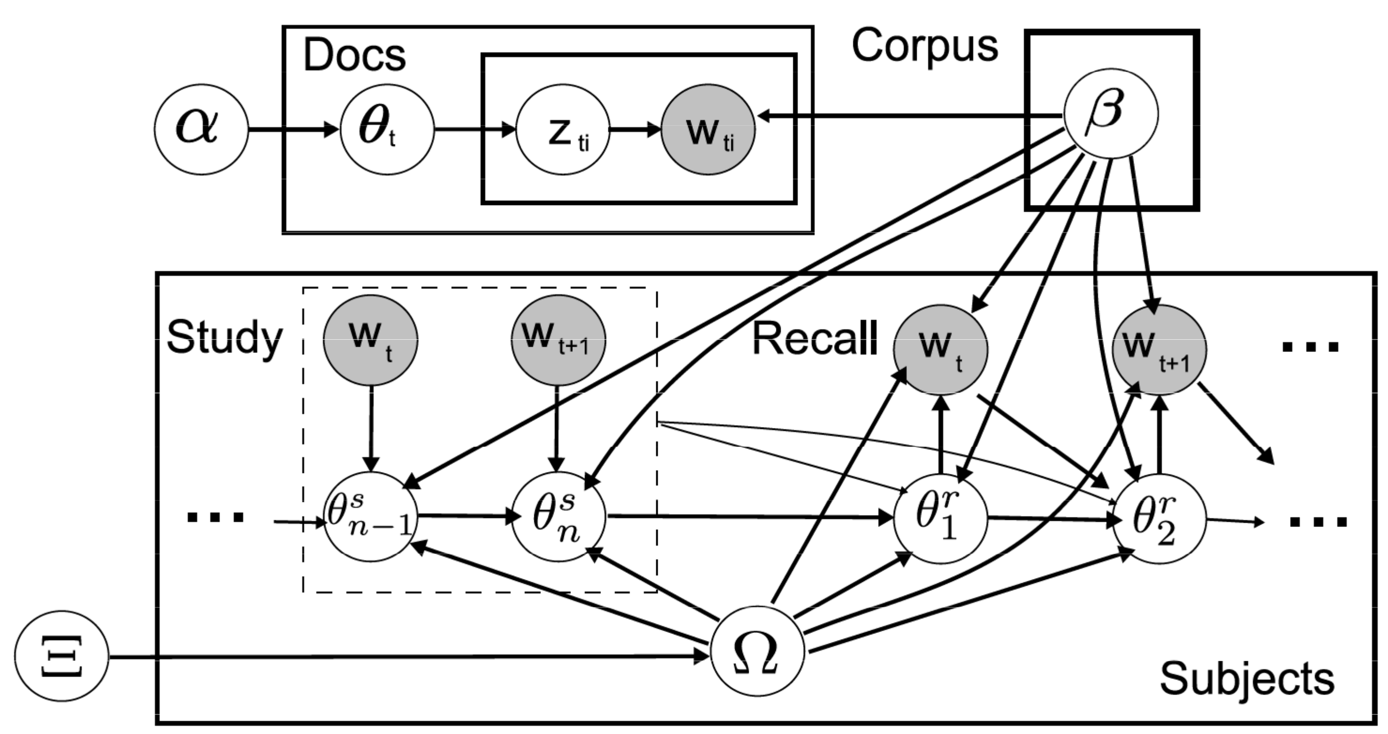
- **We develop a probabilistic model of human memory performance in free recall experiments. In these experiments, a subject first studies a list of words and then tries to recall them. We assume that memories are formed by assimilating the semantic meaning of studied words into a slowly changing latent context (represented as a distribution over topics). During recall, this context is reinstated and used as a cue for retrieving studied words.**

- Our goal is to model the psychological process of word recall in terms of a drifting mental context.
- Mental context and semantic themes in documents (e.g., from a topic model) reside in the same meaning space.
- We base our model on the Temporal Context Model (TCM) which has 3 principles, Howard and Kahana (2002):
  1. During study, words are associated with their temporal context (defined formally below)
  2. Memory retrieval involves reinstating a representation of context that was active at the time of study; and
  3. Context change is driven by features of the studied words

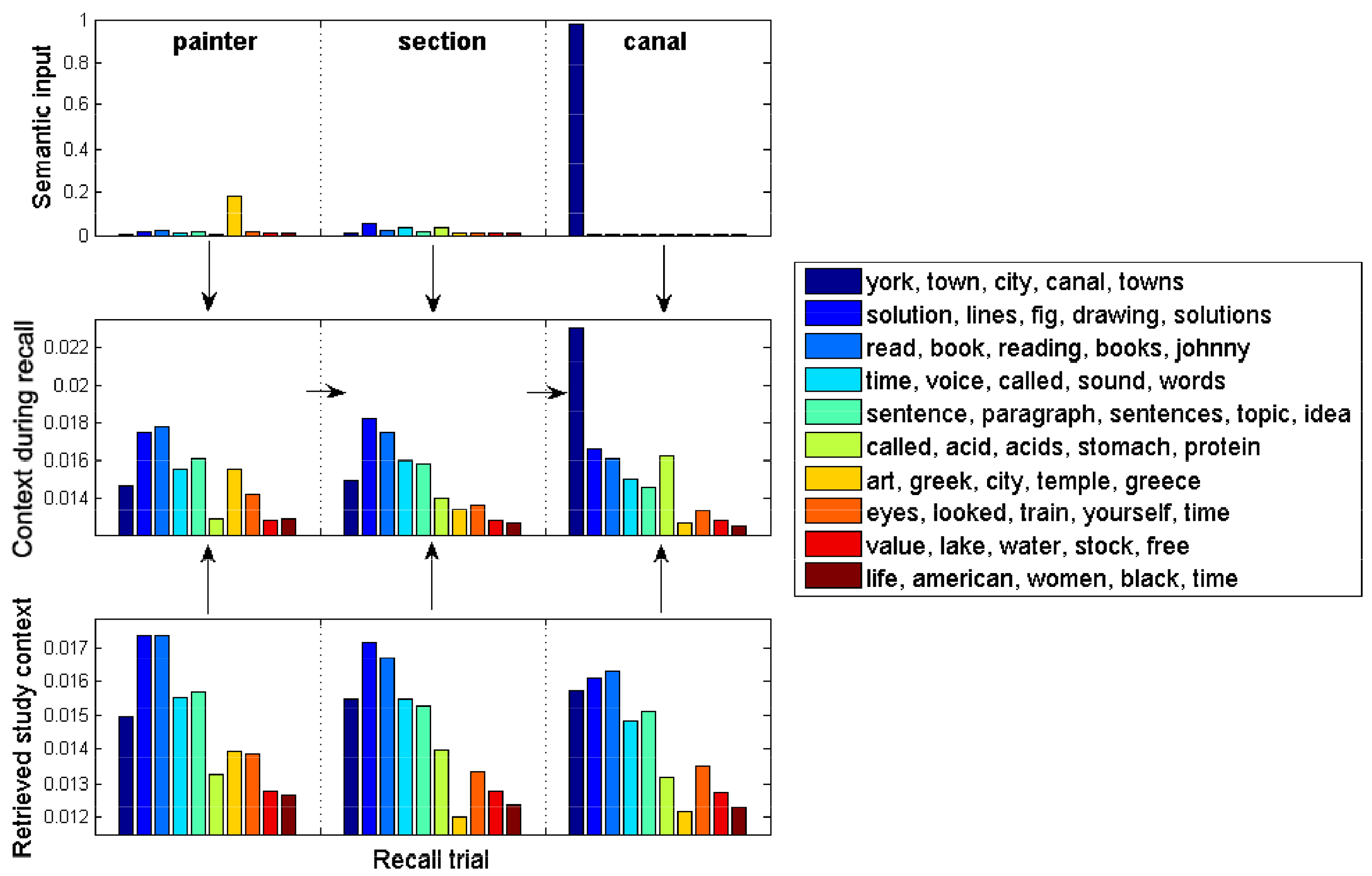
2 Modeling Temporal context and memory

- **Study Phase**
  - First, the initial mental context is drawn from a Gaussian:
  $$\theta_{s,0} \sim \mathcal{N}(\mathbf{0}, \sigma I)$$
  - Then, for each studied word the mental context drifts:
  $$\theta_{s,n} \sim \mathcal{N}(\mathbf{h}_{s,n}, \sigma I)$$
  where $\mathbf{h}_{s,n} = \underbrace{\eta_1 \theta_{s,n1}}_{\text{previous context}} + \underbrace{(1-\eta_1)\log(\tilde{\mathbf{p}}_{s,n})}_{\text{topic posterior likelihood}}$ and $\tilde{\mathbf{p}}_{s,n} \propto \beta_{\cdot,w}$
- **Recall Phase**
  - **Semantic Path:** The probability of recalling a word via the semantic path is expressed as the marginal probability of that word induced by the current context: $P_s(w) = \pi(\theta_{r,t}) \cdot \beta_{\cdot,w}$
  - **Episodic path:** The distribution over study words is expressed as a weighted sum of delta functions, where the weight for a particular study word is determined by the similarity of the context at recall to the state of context when the word was studied:
  $$P_e(w) = \frac{u_{t,w}}{\sum_i u_{t,i}}$$
  where $u_t = \sum_{n=1}^{N} \delta_{s,w_{s,n}} / d(\pi(\theta_{r,t}), \pi(\theta_{s,n}))^{\epsilon}$ and d is KL-Div
  - Each word is sampled from a mixture of the two paths:
  $w_{r,t} \sim \text{Mult}(\phi_t)$, where: $\phi_t(w) = \lambda P_s(w) + (1-\lambda) P_e(w)$
  - Finally, the context drifts according to $\theta_{r,t+1} \sim \mathcal{N}(\mathbf{h}_{r,t}, \sigma I)$
  where $\mathbf{h}_{r,t} = \underbrace{\eta_2 \theta_{r,t}}_{\text{previous context}} + \underbrace{\eta_3 \log(\tilde{\mathbf{p}}_{r,t})}_{\text{topic posterior}} + \underbrace{\eta_4 \theta_{s,n(w_{r,t})}}_{\text{retrieved context}}$
  and $n(w_{r,t})$ is a mapping from a recalled word to the index of the same word at study.

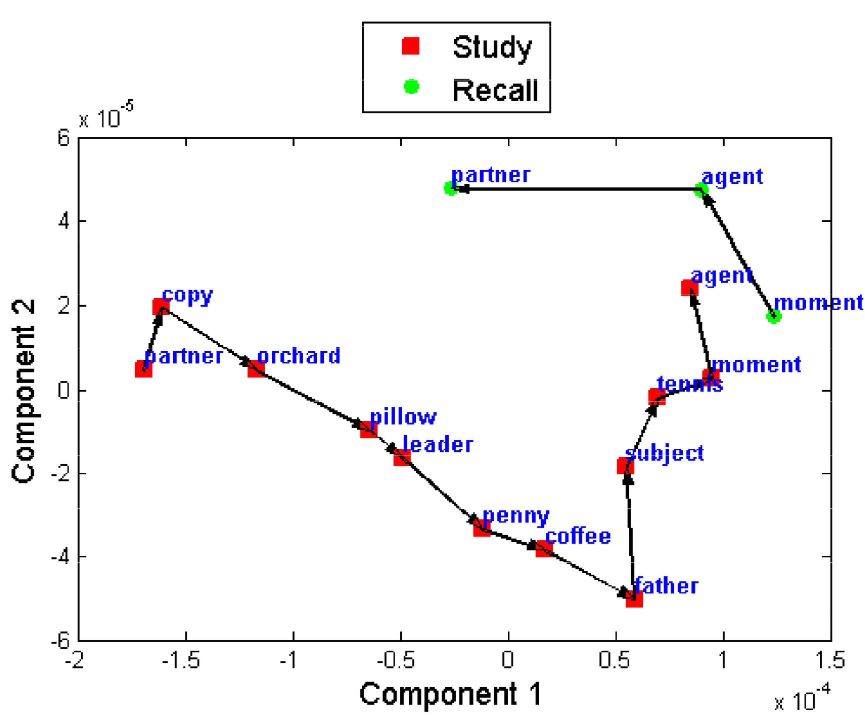
3 Graphical Temporal Context Model



4 Words influence mental context drift



Arrows schematically indicate the flow of influence between the components. The context vector at recall (Middle Row) is updated by the posterior over topics given the recalled word (Top Row) and also by retrieved study contexts (Bottom Row).

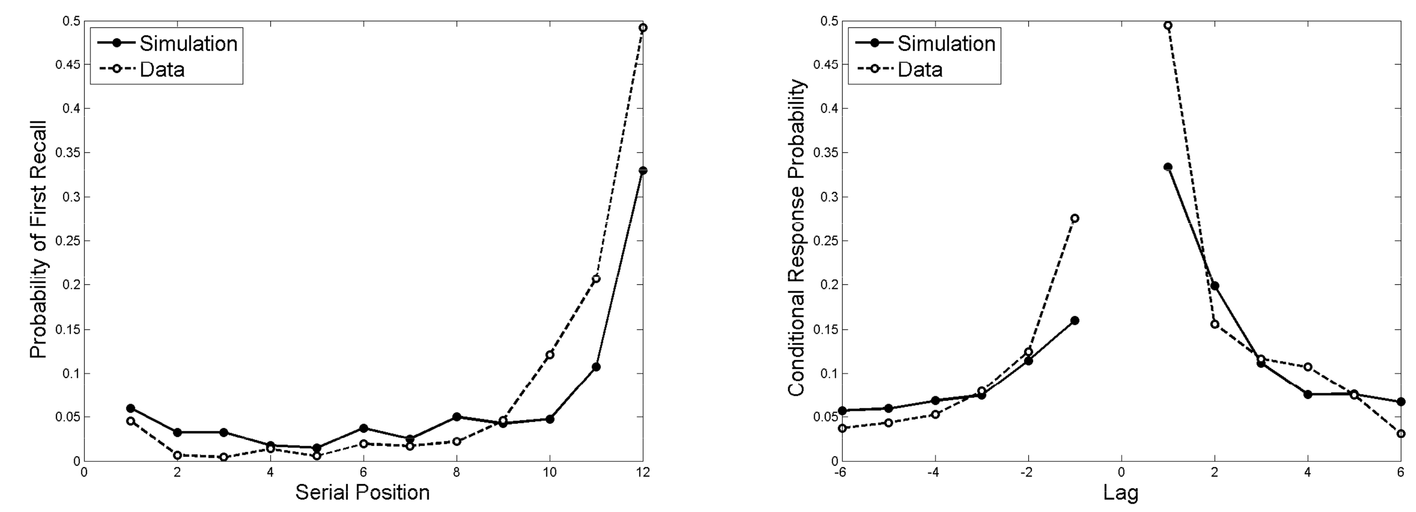


Plot of the inferred context trajectory at study and recall for a different list, in a 2-dimensional projection of the context space obtained by principal components analysis.

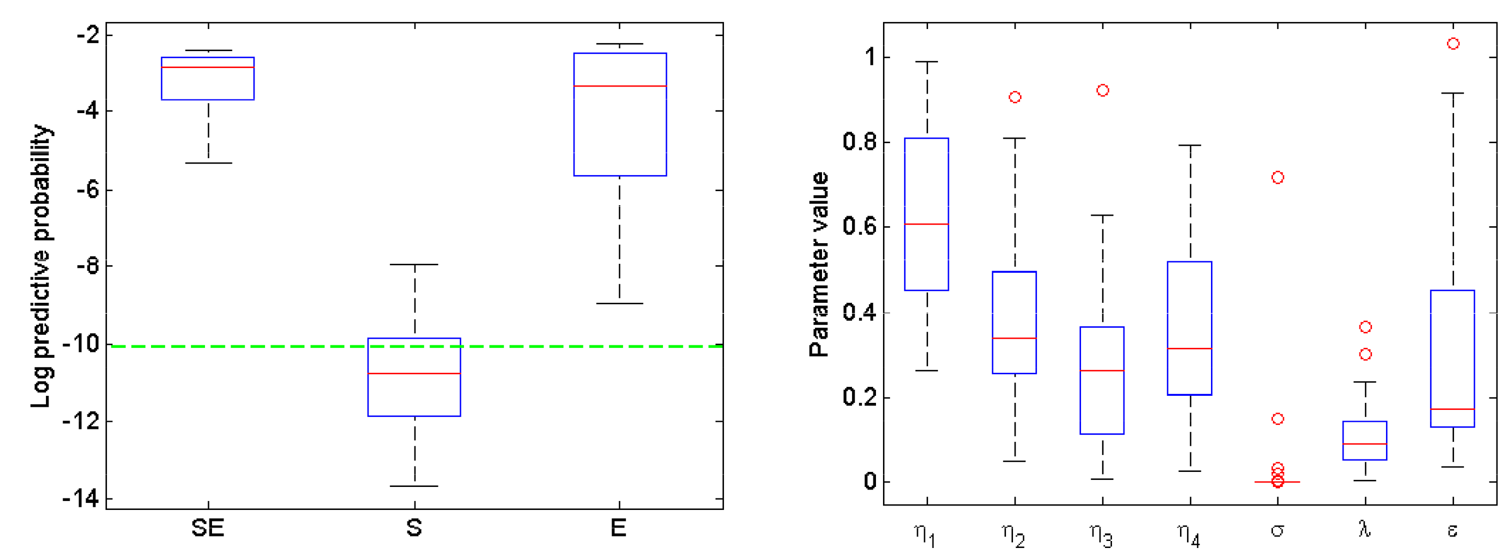
5 Inference

- Approximate intractable posterior of the parameters
$\Omega = \{\eta_{1:4}, \sigma, \lambda, \epsilon\}$ $\quad \theta = \{\theta_{s,0:N}, \theta_{r,1:T}, \Omega\}$
$$P(\theta|W) = \frac{P(w_{r,1:T}|\theta_{s,1:N}, \theta_{r,1:T}, \mathbf{w}_{s,1:N}) P(\theta_{r,1:T}|\theta_{s,1:N}) P(\theta_{s,1:N}|w_{s,1:N}, \theta_{s,0}) P(\theta_{s,0}) P(\Omega)}{P(\mathbf{w}_{s,1:N}, w_{r,1:T})}$$
- Using particle filter of (Doucet and De Freitas, 2001): for t>0
  1. Sample recall context $\theta_t^{(c)}$
  2. Compute weights $v_t^{(c)} \propto P\left(w_{r,t}|\theta_{r,t}^{(c)}\right)$.
  3. Resample the particles according to their weights.
- The posterior is approximated as: $P(\theta|\mathbf{W}) \approx \frac{1}{C}\sum_{c=1}^{C} \delta\left(\theta - \theta^{(c)}\right)$

6 Results and Conclusion



Simulated and empirical recall data. Data replotted from Howard and Kahana (1999). (Left) Probability of first recall curve. (Right) Conditional response probability curve.



(Left) Box-plot of average predictive log-probability of recalled words under different models. S: pure semantic model; E: pure episodic model. SE: Combined model. Green line indicates chance. (Right) Box-plot of inferred parameter values across subjects.

Models that use both semantic and episodic information predict word recalls better than those that use only one source.

7 Advantages of Bayesian approach

1. Easy integration with other topic models
2. Capture shared structure and individual differences across subjects
3. Can integrate other sources of data, such as brain imaging data.