

PRELIMINARIES TO AN ACCOUNT OF MULTI-PARTY CONVERSATIONAL TURN-TAKING AS AN ANTIFERROMAGNETIC SPIN GLASS

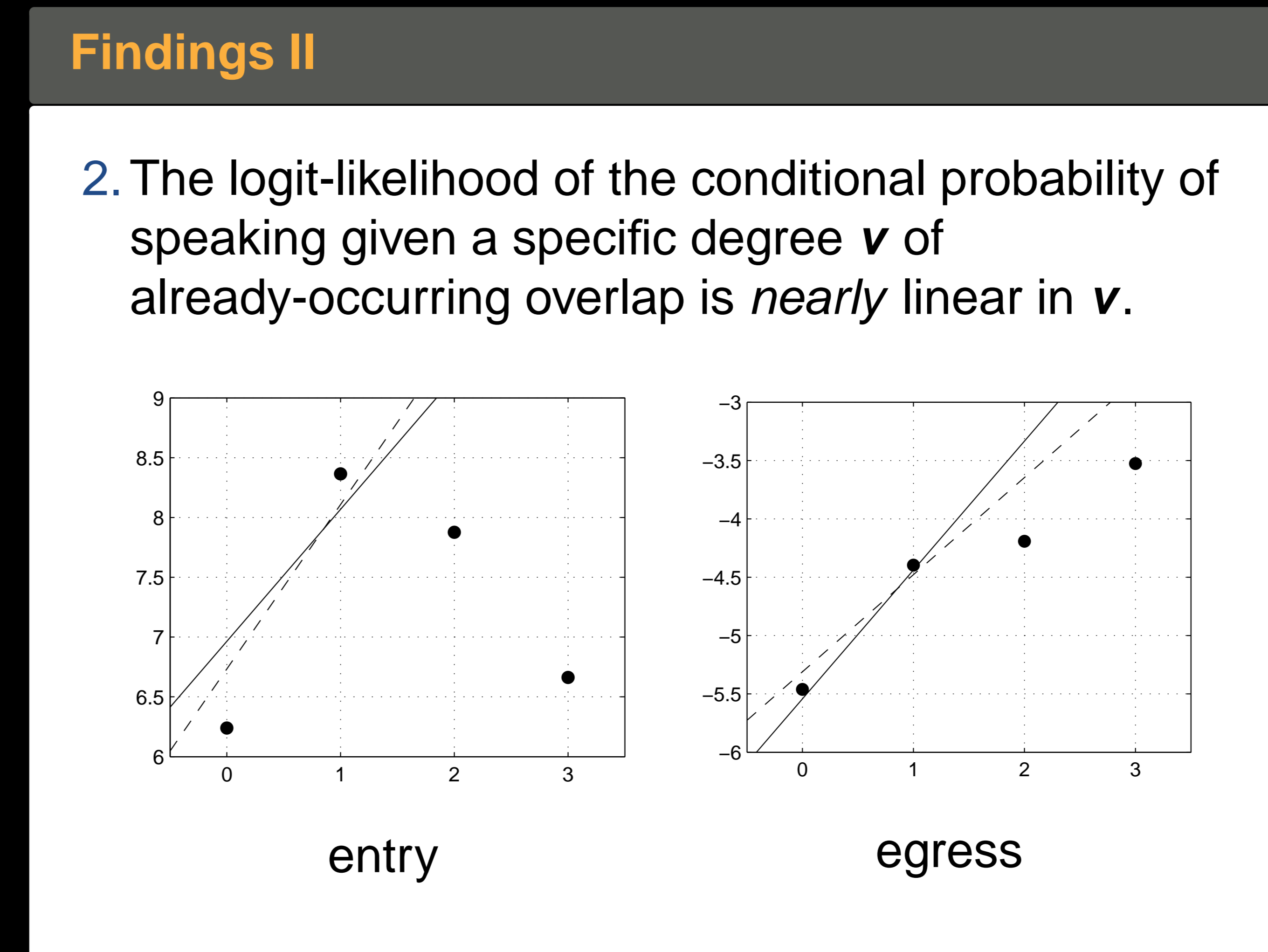
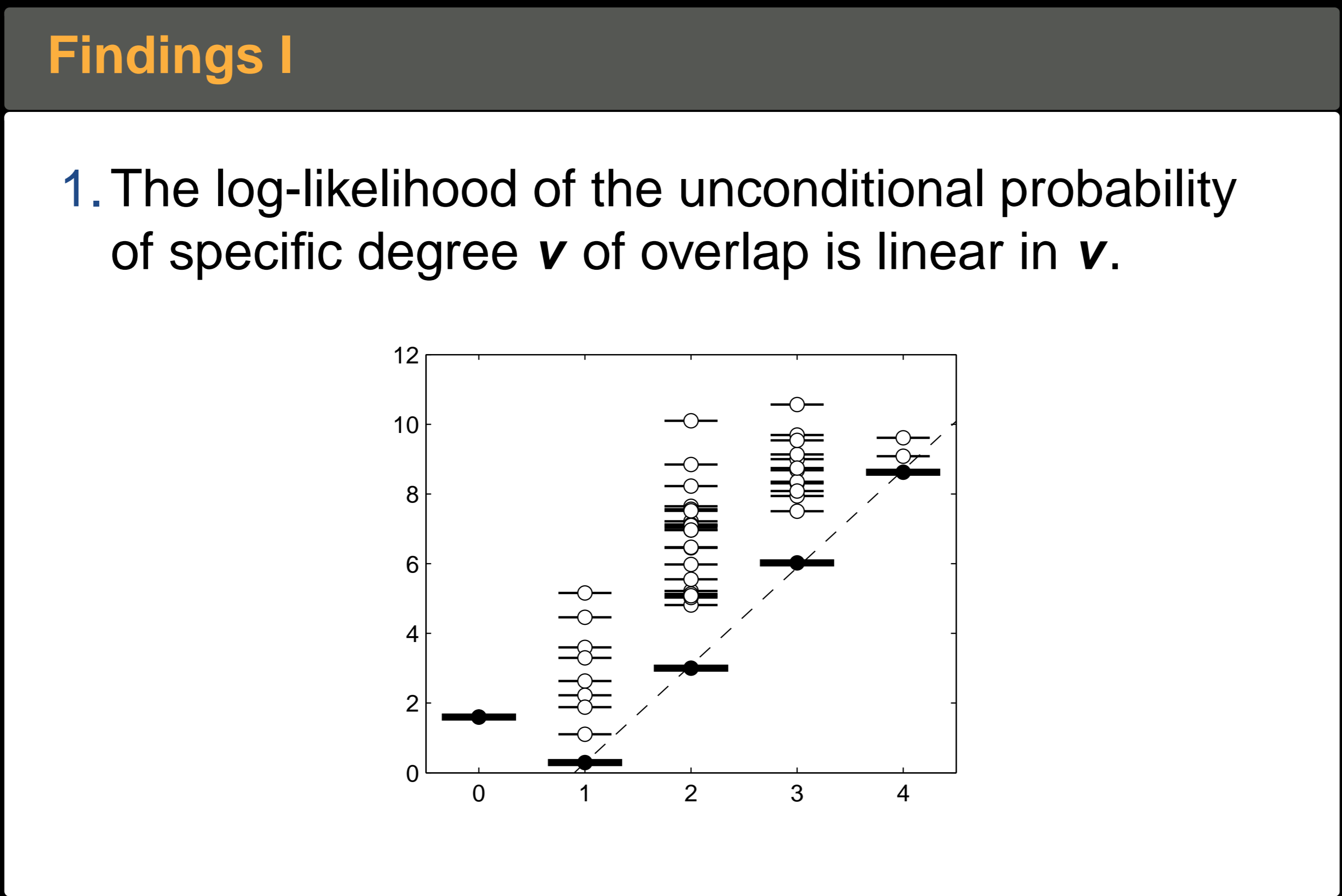
Kornel Laskowski, Mattias Heldner & Jens Edlund
KTH Speech Music and Hearing, Stockholm, Sweden



Goal

Argue for a **parametric form** which reduces the complexity of a model Θ of **multi-party turn-taking**.

The role of Θ is to provide the likelihood of an arbitrary-size **speech/non-speech chronogram**

$$Q = \begin{bmatrix} \square & \square & \square & \square & \square & \square & \blacksquare & \blacksquare & \square & \square & \square & \square \\ \blacksquare & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \dots & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \dots \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{bmatrix}$$


Potential Impact

I. Significant reduction in number of parameters:
 Conversation-specific form:
 $K (2^{mK}) \mapsto K + mK^2$
 Conversation-independent form:
 $(2K)^m \mapsto 1 + 2m$

II. The parametric model is a neural network:
 Predictions can easily be extended to rely on *any* feature type at instant $t - 1$.

III. Model structure is directly interpretable.
 Pairwise influence overtly coded in the magnitude of the off-diagonal entries of W .

Standard Non-Parametric Form

Factor in time, Markovian assumption of order $m = 1$:

$$P \left(\begin{array}{cccccccc} \square & \square & \square & \square & \square & \square & \blacksquare & \blacksquare & \square & \square & \square & \square \\ \blacksquare & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \dots & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \dots \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} \right) = \dots \times P \left(\begin{array}{c|c} \square & \square \\ \blacksquare & \square \\ \square & \square \\ \square & \square \\ \square & \square \end{array} \right) \times P \left(\begin{array}{c|c} \square & \square \\ \square & \square \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array} \right) \times P \left(\begin{array}{c|c} \blacksquare & \square \\ \square & \square \\ \square & \square \\ \square & \square \\ \square & \square \end{array} \right) \times \dots$$

Factor each factor, assuming participants are *conditionally independent*:

$$P \left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array} \right) = P_1 \left(\begin{array}{c|c} 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \square \\ 5 & \square \end{array} \right) \times P_2 \left(\begin{array}{c|c} 1 & \square \\ 2 & \blacksquare \\ 3 & \square \\ 4 & \square \\ 5 & \square \end{array} \right) \times P_3 \left(\begin{array}{c|c} 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \blacksquare \\ 5 & \square \end{array} \right) \times P_4 \left(\begin{array}{c|c} 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \square \\ 5 & \blacksquare \end{array} \right) \times P_5 \left(\begin{array}{c|c} 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \square \\ 5 & \square \end{array} \right)$$

This is the **conversation-specific** model. It has K submodels, each of $2 \cdot 2^K$ parameters $\rightarrow K \cdot (2^K)^m$ degrees of freedom in total.

For the **conversation-independent** model,

$$P \left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array} \right) = P \left(\begin{array}{c|c} 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \square \\ 5 & \square \end{array} \right) \times P \left(\begin{array}{c|c} 2 & \square \\ 1 & \square \\ 3 & \square \\ 4 & \square \\ 5 & \square \end{array} \right) \times P \left(\begin{array}{c|c} 3 & \square \\ 1 & \square \\ 2 & \square \\ 4 & \square \\ 5 & \square \end{array} \right) \times P \left(\begin{array}{c|c} 4 & \blacksquare \\ 1 & \square \\ 2 & \square \\ 3 & \square \\ 5 & \square \end{array} \right) \times P \left(\begin{array}{c|c} 5 & \square \\ 1 & \square \\ 2 & \square \\ 3 & \square \\ 4 & \blacksquare \end{array} \right)$$

$$= P \left(\begin{array}{c|c} 1 & \square \\ oth & 2 \end{array} \right) \times P \left(\begin{array}{c|c} 2 & \square \\ oth & 1 \end{array} \right) \times P \left(\begin{array}{c|c} 3 & \square \\ oth & 2 \end{array} \right) \times P \left(\begin{array}{c|c} 4 & \blacksquare \\ oth & 1 \end{array} \right) \times P \left(\begin{array}{c|c} 5 & \square \\ oth & 2 \end{array} \right)$$

There is only 1 submodel, of $2 \cdot (2 \cdot K)$ parameters $\rightarrow (2 \cdot K)^m$ degrees of freedom in total.

A "Natural" Consequence of the Logistic Fit

Logistic regression \equiv Glauber dynamics \equiv most prosaic feed-forward neural network structure:

$$P_k (q_t[k] = \blacksquare | q_{t-1}) = \frac{1}{1 + e^{-\beta h_{t-1}[k]}}$$

$$P_k (q_t[k] = \square | q_{t-1}) = 1 - P_k (q_t[k] = \blacksquare | q_{t-1})$$

$$h_{t-1} = \mathbf{b} + \mathbf{W} \cdot \mathbf{q}_{t-1} \quad \text{with } \{\square, \blacksquare\} \equiv \{0, 1\}$$

Conversation-specific model:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} W_{11} & W_{12} & W_{13} & W_{14} & W_{15} \\ W_{21} & W_{22} & W_{23} & W_{24} & W_{25} \\ W_{31} & W_{32} & W_{33} & W_{34} & W_{35} \\ W_{41} & W_{42} & W_{43} & W_{44} & W_{45} \\ W_{51} & W_{52} & W_{53} & W_{54} & W_{55} \end{bmatrix}$$

$K + mK^2$ parameters in total (and degrees of freedom)

This happens to be an **infinite-range anti-ferromagnetic spin glass (with slow connections)**.

Conversation-independent model:

$$\mathbf{b} = \begin{bmatrix} b \\ b \\ b \\ b \\ b \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} W_+ & W_- & W_- & W_- & W_- \\ W_- & W_+ & W_- & W_- & W_- \\ W_- & W_- & W_+ & W_- & W_- \\ W_- & W_- & W_- & W_+ & W_- \\ W_- & W_- & W_- & W_- & W_+ \end{bmatrix}$$

$1 + 2m$ parameters in total (and degrees of freedom)

This happens to be an **infinite-range Ising anti-ferromagnet with slow connections**.