An Associative Model of Inference in Statistical Word Learning

George Kachergis, Chen Yu, & Richard M. Shiffrin

Indiana University

July 18, 2011

Math Psych 2011
Goal

• Demonstrate that a model operating on two *simple associative mechanisms* captures a variety of human *word-learning* behavior, including *inference* and *bootstrapping*.
Learn which words go with which objects.
“stigson”

“manu”
“bosa”

“stigson”
“manu”
“stigson”
“bosa”
“manu”

“stigson”

“bosa”
“manu”

“stigson”

“bosa”

1-to-2 mapping?
1-to-3 mapping?
## Word-Object Co-occurrences

<table>
<thead>
<tr>
<th>words</th>
<th>objects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>![Image 360x338 to 396x376]</td>
</tr>
<tr>
<td><strong>bosa</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>stigson</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>manu</strong></td>
<td>1</td>
</tr>
</tbody>
</table>
Mutual Exclusivity (ME)

• Learners try to map words to referents in a 1-to-1 way (Markman, 1990)
• Drastically reduces learning complexity:
  – don’t have to consider all possible mappings
• What about synonyms and homonyms?
  – ME must be flexible—not strict
• Whence the ME bias?
# Early Co-occurrence Matrix

<table>
<thead>
<tr>
<th>objects</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

**Cross-situational learning:**
Each pair (1-6) appears with 3 other pairs

- 9 trials
- 2 pairs/trial

Kachergis, Yu, & Shiffrin, 2009
## Early+Late Co-occurrence Matrix

<table>
<thead>
<tr>
<th>words</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

**ME-based inference:**
Each late pair (7-12) appears with only one early pair (e.g., 7 with 1)

3 Early, 6 Late Re�etitions Condition

Kachergis, Yu, & Shiffrin, 2009
Co-occurrence Matrix (3 Early, 6 Late)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

What is learned?

Within-stage:

\(w_1-o_1?\)
\(w_7-o_7?\)

Across-stage:

\(w_1-o_7?\)
\(w_7-o_1?\)

Kachergis, Yu, & Shiffrin, 2009
Co-occurrence Matrix (3 Early, 6 Late)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Each word (e.g., 1) tested twice with 11AFC:
- Once without its early object (1)
- Once without its late object (7)

What is learned?

Within-stage:
- \( w_1-o_1 \) early
- \( w_7-o_7 \) late

Across-stage:
- \( w_1-o_7 \)?
- \( w_7-o_1 \)?

Kachergis, Yu, & Shiffrin, 2009
Varying Evidence Strength

- Early stage:
  - 0, 3, 6, or 9 repetitions/pair (Within-subjects)
- Late stage:
  - 3, 6, or 9 repetitions/pair (Between-subjects)
- Do they infer $w_7-o_7$ using $w_1-o_1$? Are cross-stage pairings learned?
- Does learning depend on the number/ratio of late and early repetitions?

Kachergis, Yu, & Shiffrin, 2009
Inference Results

Early pairs ($w_1-o_1$) learned well, and late pairs ($w_7-o_7$) inferred as well.

Cross-stage pairs learned above chance.

No effect of early repetitions (3,6,9).
Inference Results

Increased learning of cross-stage pairs.

Early $W_1 \rightarrow O_1$

Late $W_7 \rightarrow O_7$
Inference Results

- Same strong inference of late from early
- Weak across-stage learning—but some!
- No effect of early evidence—3 as good as 9
- Increasing learning with more late reps
Representation & Mechanisms

• For the \( m \) words and objects on each trial, distribute a **fixed** amount (\( \chi \)) of associative weight among the possible pairings

• Attention distribution functions:
  – Unbiased: equal portion given to each association
  – Prior Knowledge Bias: proportional to current association strength

\[
M_{w,o} = M_{w,o} + \frac{M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} M_{w,o}}
\]
Word-Object Associations

\[
M = \begin{array}{ccc}
\text{objects} \\
\text{words} \\
\hline
\text{bosa} & \text{image} & \text{image} \\
\text{stigson} & \text{image} & \text{image} \\
\text{manu} & \text{image} & \text{image} \\
\end{array}
\]

Initially empty
Unbiased vs. Prior Knowledge Bias

**Trial 0:**

\[ \chi = 1 \]

<table>
<thead>
<tr>
<th></th>
<th>bosa</th>
<th>stigson</th>
<th>manu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="bosa.png" alt="Image" /></td>
<td><img src="stigson.png" alt="Image" /></td>
<td><img src="manu.png" alt="Image" /></td>
</tr>
</tbody>
</table>

\[ M_{w,o} = M_{w,o} + \frac{\chi}{|S|^2} \]

\[ M_{w,o} = M_{w,o} + \frac{M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} M_{w,o}} \]
Unbiased vs. Prior Knowledge Bias

Trial 1:

<table>
<thead>
<tr>
<th></th>
<th>stigson</th>
<th>manu</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stigson</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>manu</td>
<td>.25</td>
<td>.25</td>
</tr>
</tbody>
</table>

\[
M_{w,o} = M_{w,o} + \frac{\chi}{|S|^2}
\]

\[
M_{w,o} = M_{w,o} + \frac{M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} M_{w,o}}
\]
## Unbiased vs. Prior Knowledge Bias

### Trial 2:

<table>
<thead>
<tr>
<th></th>
<th>bosa</th>
<th>stigson</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td>.25</td>
<td>.04</td>
</tr>
<tr>
<td>stigson</td>
<td>.25</td>
<td>1.14</td>
</tr>
<tr>
<td>manu</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

\[
M_{w,o} = M_{w,o} + \frac{\chi}{|S|^2}
\]

\[
M_{w,o} = M_{w,o} + \frac{M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} M_{w,o}}
\]
Unbiased vs. Prior Knowledge Bias

Test:
Hear each word, pick an object

\[ p(o \mid w) = \]

\begin{array}{|c|c|c|}
\hline
& \text{bosa} & \text{stigson} & \text{manu} \\
\hline
\text{bosa} & .49 & .49 & .02 \\
\text{stigson} & .25 & .50 & .25 \\
\text{manu} & .02 & .49 & .49 \\
\hline
\end{array}

\begin{array}{|c|c|c|}
\hline
& \text{bosa} & \text{stigson} & \text{manu} \\
\hline
\text{bosa} & .44 & .44 & .11 \\
\text{stigson} & .03 & .80 & .17 \\
\text{manu} & .02 & .49 & .49 \\
\hline
\end{array}

\[ M_{w,o} = M_{w,o} + \frac{\chi}{|S|^2} \]
\[ M_{w,o} = M_{w,o} + \frac{M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} M_{w,o}} \]
Prior Knowledge Bias

Looks good!

...wait: $w_7-o_7$
Prior Knowledge vs. Uncertainty

• Trial 1: everything is *novel* and we should associate everything equally
• On trials with *new* and *old* pairs, how should we distribute attention?
• An *old w* may go with an *old o*, and $M_{w,o}$ can be increased in proportion to current strength
• But a *novel w* and *novel o* may be likely to go together, and should be strongly associated
• A *novel w* should not be linked to a *known o*
Uncertainty = Entropy

• Learners know about what words and objects they know (or don’t know)

• Entropy is a measure of the uncertainty about a random variable (e.g., a stimuli’s associations)

\[
H(w) = - \sum_{o \in M} p(o|w) \cdot \log(p(o|w))
\]

e.g., \( w_1 = [1, 0, 0, 0, 0, 0] \) \( w_2 = [.2, .2, .2, .2, .2] \)

\( H(w_1) = 0 \) \( H(w_2) = 2.32 \)
PK & Entropy Bias Model

- Learn more about **uncertain** words and objects—especially pairings with both—but also strengthen **strong** associations.

\[ M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}} \]
PK & Entropy Bias Model

- Learn more about uncertain words and objects—especially pairings with both—but also strengthen strong associations
- Scaling parameter ($\lambda$) to change relative weight of uncertainty and PK biases

$$M_{w,o} = M_{w,o} + \frac{e^{\lambda \cdot (H(w)+H(o))} \cdot (M_{w,o} \cdot \chi)}{\sum_{w \in S} \sum_{o \in S} e^{\lambda \cdot (H(w)+H(o))} \cdot M_{w,o}}$$
PK & Uncertainty Bias Model

**Trial 1:**

<table>
<thead>
<tr>
<th></th>
<th>stigson</th>
<th>manu</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stigson</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>manu</td>
<td>.01</td>
<td>.01</td>
</tr>
</tbody>
</table>

\[ H(w) = H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi \]

\[ M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}} \]

\[ \chi = 1 \]

\[ \lambda = 1 \]

Update experienced stimuli
PK & Uncertainty Bias Model

Trial 1:  
- stigson
- manu

<table>
<thead>
<tr>
<th></th>
<th>H(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

Distribute attention based on competing PK and uncertainty biases—equal in the beginning

\[ M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}} \]
PK & Uncertainty Bias Model

**Trial 2:**

<table>
<thead>
<tr>
<th></th>
<th>bosa</th>
<th>stigson</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>stigson</td>
<td>.01</td>
<td>.26</td>
</tr>
<tr>
<td>manu</td>
<td>.01</td>
<td>.26</td>
</tr>
</tbody>
</table>

\[
M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}}
\]

- **H(w)**
  - bosa: 1.58
  - stigson: 1.12
  - manu: 1.12

- **H(o)**
  - bosa: 1.58
  - stigson: 1.12
  - manu: 1.12

**Uncertainty**  **Prior Knowledge**
PK & Uncertainty Bias Model

**Trial 2:**
- **bosa**
- **stigson**

<table>
<thead>
<tr>
<th></th>
<th>bosa</th>
<th>stigson</th>
<th>H(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td>.09</td>
<td>.01</td>
<td>1.25</td>
</tr>
<tr>
<td>stigson</td>
<td>.06</td>
<td>.26</td>
<td>0.93</td>
</tr>
<tr>
<td>manu</td>
<td>.01</td>
<td>.26</td>
<td>1.12</td>
</tr>
</tbody>
</table>

\[ H(o) = 1.25 \quad 0.93 \quad 1.12 \]

\[ M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}} \]
Test:
Hear each word,
pick an object

\[
p(o \mid w) = \]

<table>
<thead>
<tr>
<th></th>
<th>🍎</th>
<th>🍏</th>
<th>🍓</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td>.56</td>
<td>.38</td>
<td>.06</td>
</tr>
<tr>
<td>stigson</td>
<td>.04</td>
<td>.77</td>
<td>.19</td>
</tr>
<tr>
<td>manu</td>
<td>.02</td>
<td>.49</td>
<td>.49</td>
</tr>
</tbody>
</table>

\[
M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}}
\]
PK & Uncertainty Bias Model

**Test:**
Hear each word, pick an object

\[ p(o \mid w) = \]

<table>
<thead>
<tr>
<th></th>
<th>![Image 1]</th>
<th>![Image 2]</th>
<th>![Image 3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>bosa</td>
<td>.56</td>
<td>.38</td>
<td>.06</td>
</tr>
<tr>
<td>stigson</td>
<td>.04</td>
<td>.77</td>
<td>.19</td>
</tr>
<tr>
<td>manu</td>
<td>.02</td>
<td>.49</td>
<td>.49</td>
</tr>
</tbody>
</table>

- It ‘infers’ bosa-![Image 4]!
- Still doesn’t *know* manu (retrospective re-evaluation)

\[
M_{w,o} = M_{w,o} + \frac{H(w) \cdot H(o) \cdot M_{w,o} \cdot \chi}{\sum_{w \in S} \sum_{o \in S} H(w) \cdot H(o) \cdot M_{w,o}}
\]
Adaptive Inference: Model Results

MSE = 0.005
Varied Pair Frequency

Pairs of each frequency only appear with pairs of that frequency.

3 pairs/trial
36 training trials

Kachergis, Yu, & Shiffrin 2009
Varied Pair Frequency

Clear, pure frequency effect.

N=36

Kachergis, Yu, & Shiffrin 2009
Varied Frequency & CD

Pairs of all frequencies co-occur randomly

Kachergis, Yu, & Shiffrin 2009
Frequency & Contextual Diversity

• The more diverse a pair’s context, the easier it is to learn—context diversity.

• Confounded with freq: high frequency pairs will co-occur more often with other pairs.
Clear, pure frequency effect in Low CD condition.

In High CD condition, low and medium frequency pairs are boosted:

Mixing high freq items with low freq items helps learning.
Bootstrapping

• When HF pairs appear with LF pairs, the learning of LF pairs is boosted
• Likely due to *bootstrap* (i.e., use knowledge of HF pairs attained early in training to constrain later pair learning—*assuming ME*)
• Does the Prior Knowledge and Uncertainty Bias model bootstrap?
Varied Frequency & CD Results

MSE=.004
Varied Frequency & CD Results

Model learning trajectories:

Exp : Low CD

Exp : High CD

MSE=.004
Summary

• Associative learning model that assumes learners know what they know, and how well they know it
• A fixed learning rate per trial representing limited attention capacity
• Competing uncertainty and prior knowledge biases yields inference and bootstrapping
• Matches many empirical human data
• Explains mutual exclusivity bias
Future Directions

• The model looks like a Bayesian update rule, combining prior strength and uncertainty (but doesn’t update beliefs for stimuli not on a given trial)—explore this relationship

• The model also exhibits some associative learning effects (e.g., blocking)—how far can it go? In word learning, are words outcomes or cues?
Thank you!

Questions?

gkacherg@indiana.edu

Special thanks to Greg Cox, Stephen Denton, and John Kruschke
# ME Investigations

<table>
<thead>
<tr>
<th>Training Stage</th>
<th>ME (Yurovsky &amp; Yu, 2008)</th>
<th>ME (Ichinco, et al., 2009)</th>
<th>Blocking (Kamin, 1968)</th>
<th>Present study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>$w_1-o_1$</td>
<td>$w_1-o_1$</td>
<td>$q_1-o$</td>
<td>$w_1-o_1$</td>
</tr>
<tr>
<td>Late</td>
<td>$w_1-o_2$ (or $w_2-o_1$)</td>
<td>$w_1-{o_1, o_2}$</td>
<td>${q_1, q_2}-o$</td>
<td>${w_1, w_2}-{o_1, o_2}$</td>
</tr>
<tr>
<td>Test</td>
<td>$w_1$-? (prefer early, can learn both)</td>
<td>$w_1$-? (prefer early, rarely learn both)</td>
<td>Don’t learn $q_2-o_1$</td>
<td>$w_1$-? $w_2$-?</td>
</tr>
</tbody>
</table>
Basic Co-occurrence Matrix Math

- $f_i =$ frequency of item $i$ (i.e., diagonal)
- $c =$ context size (# of items per trial)
- $n_i =$ # of distinct items $i$ can appear with
  (i.e., vocabulary size)
- Expected value of an off-diagonal cell:
  \[
  \frac{f(c-1)}{n-1}
  \]
  e.g., $f=6$, $c=4$, $n=18$: E.V. = 1.06
Varied Frequency & CD Results

CD & Frequency Exp. 3

Proportion Correct

Condition

Low CD  High CD  Med CD (3/6)  Med CD (3/9)

N=36  N=31

Kachergis, Yu, & Shiffrin 2009
Varied Frequency & CD Results
Mutual Exclusivity

• Violate ME if learn:

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]

  or

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]

  or

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]

  or

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]

• Follow ME if learn:

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]

  or

  \[\begin{align*}
  w_1 &\rightarrow o_1 \\
  w_7 &\rightarrow o_7 \\
  \end{align*}\]
ME Compliant

Learning of Mutual Exclusivity-respecting Pairings

Proportion of Pairings Learned

Early Repetitions
- 0
- 3
- 6
- 9

Late Pair Repetitions
- 3
- 6
- 9
ME Violations

Learning of Mutual Exclusivity-violating Pairings

- Early Repetitions
- Late Pair Repetitions

Proportion of Pairings Learned

0, 3, 6, 9 repetitions