Analyzing human feature learning as non-parametric Bayesian inference

By Joseph L. Austerweil and Thomas L. Griffiths
Department of Psychology, UC Berkeley
Features
Features

- Features are the elementary units that objects are built from.
Features

- Features are the elementary units that objects are built from.

- They represent the commonalities and differences of objects, which can be used for:
  - e.g., object recognition
Feature Change
Feature Change

- *Unitization* - When two or more distinct features combine into a single feature
Feature Change

- *Unitization* - When two or more distinct features combine into a single feature

- *Differentiation* - When a new feature is a (strict) subset of a previous feature
Unitization

Shiffrin & Lightfoot (1997)
Unitization

- Goal: unitize original features by repeated exposure

Shiffrin & Lightfoot (1997)
Unitization

- Goal: unitize original features by repeated exposure
- Task: visual search where single features cannot aid search

1 object was target and the task was to find it in a scene where the other 3 objects were distractors

Shiffrin & Lightfoot (1997)
Unitization

- **Goal:** unitize original features by repeated exposure
- **Task:** visual search where single features cannot aid search

1 object was target and the task was to find it in a scene where the other 3 objects were distractors

- Need 2 features to discriminate between objects

Shiffrin & Lightfoot (1997)
Unitization

• Goal: unitize original features by repeated exposure

• Task: visual search where single features cannot aid search

1 object was target and the task was to find it in a scene where the other 3 objects were distractors

• Need 2 features to discriminate between objects

Shiffrin & Lightfoot (1997)
Unitization

- **Goal**: unitize original features by repeated exposure
- **Task**: visual search where single features cannot aid search

1 object was target and the task was to find it in a scene where the other 3 objects were distractors

Need 2 features to discriminate between objects

Shiffrin & Lightfoot (1997)
Unitization

$X_1 \quad X_2 \quad X_3 \quad X_4$
Unitization

```
x1 1 1 1 0 0 0 0
x2 0 1 0 1 0 1 0
x3 0 0 1 1 1 1 0
x4 1 0 0 0 0 1 1
```

feature ownership matrix
Unitization

$x_1$

$x_2$

$x_3$

$x_4$

feature ownership matrix
Unitization

feature ownership matrix
Unitization

feature ownership matrix
Unitization

X1

X2

X3

X4

feature ownership matrix

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>X3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>X4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Unitization

feature ownership matrix
Unitization

Perceptual Learning (unitize)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$x_4$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

feature ownership matrix
Unitization

Perceptual Learning (unitize)

<table>
<thead>
<tr>
<th></th>
<th>Feature Ownership Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>x_2</td>
<td>0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td>x_3</td>
<td>0 0 1 1 1 1 1 0</td>
</tr>
<tr>
<td>x_4</td>
<td>1 0 0 0 0 1 1 1</td>
</tr>
</tbody>
</table>

Perceptual Learning (unitize)
Differentiation

Pevtzow & Goldstone (1994)
Differentiation

- Goal: differentiate by repeated categorizations
Differentiation

- **Goal:** differentiate by repeated categorizations
- **Task:** show object, what category is it in? (Feedback given)

Pevtzow & Goldstone (1994)
Differentiation

- Goal: differentiate by repeated categorizations
- Task: show object, what category is it in? (Feedback given)
- Two Training Groups:

Pevtzow & Goldstone (1994)
Differentiation

• Goal: differentiate by repeated categorizations

• Task: show object, what category is it in? (Feedback given)

• Two Training Groups:
  • horizontal
cat 1 = \{A,B\}, cat 2 = \{C,D\}

Pevtzow & Goldstone (1994)
Differentiation

- **Goal:** differentiate by repeated categorizations
- **Task:** show object, what category is it in? (Feedback given)
- **Two Training Groups:**
  - **horizontal**
    - cat 1 = \{A,B\}, cat 2 = \{C,D\}
  - **vertical**
    - cat 1 = \{A,C\}, cat 2 = \{B,D\}

Pevtzow & Goldstone (1994)
Differentiation

• Goal: differentiate by repeated categorizations

• Task: show object, what category is it in? (Feedback given)

• Two Training Groups:
  • horizontal
cat 1 = \{A,B\}, cat 2 = \{C,D\}
  • vertical
cat 1 = \{A,C\}, cat 2 = \{B,D\}

• Participants learn features appropriate to their training condition

Pevtzow & Goldstone (1994)
Differentiation

- Goal: differentiate by repeated categorizations
- Task: show object, what category is it in? (Feedback given)
- Two Training Groups:
  - horizontal
cat 1 = \{A,B\}, cat 2 = \{C,D\}
  - vertical
cat 1 = \{A,C\}, cat 2 = \{B,D\}
- Participants learn features appropriate to their training condition

Pevtzow & Goldstone (1994)
Differentiation

- Goal: differentiate by repeated categorizations
- Task: show object, what category is it in? (Feedback given)
- Two Training Groups:
  - horizontal
    cat 1 = {A, B}, cat 2 = {C, D}
  - vertical
    cat 1 = {A, C}, cat 2 = {B, D}
- Participants learn features appropriate to their training condition

Pevtzow & Goldstone (1994)
Formalizing Feature Representation Inference

Austerweil & Griffiths (in submission)
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X)$$
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$
Formalizing Feature Representation Inference

• Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_{Z} P(Z|X) = \arg \max_{Z} P(X|Z)P(Z)$$

• $P(Z)$ = prior on feature ownership matrices
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg\max_Z P(Z|X) = \arg\max_Z P(X|Z)P(Z)$$

- $P(Z)$ = prior on feature ownership matrices

Austerweil & Griffiths (in submission)
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

- $P(Z) =$ prior on feature ownership matrices

- The Indian Buffet Process is a “good” prior on feature ownership matrices. (Griffiths & Ghahramani 2006)
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$
  
  $$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

- $P(Z)$ = prior on feature ownership matrices

- The Indian Buffet Process is a “good” prior on feature ownership matrices. (Griffiths & Ghahramani 2006)

- Multiple features per object and unfixed # of features
Formalizing Feature Representation Inference

• Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

• $P(Z) = \text{prior on feature ownership matrices}$

• The Indian Buffet Process is a “good” prior on feature ownership matrices. (Griffiths & Ghahramani 2006)

• Multiple features per object and unfixed # of features

• $P(X|Z) = \text{likelihood of objects given their features}$

Austerweil & Griffiths (in submission)
Formalizing Feature Representation Inference

- Goal: form the “best” feature representation, $\hat{Z}$, for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

- $P(Z) =$ prior on feature ownership matrices

- The Indian Buffet Process is a “good” prior on feature ownership matrices. (Griffiths & Ghahramani 2006)

- Multiple features per object and unfixed # of features

- $P(X|Z) =$ likelihood of objects given their features

- Combines features to form objects (“noisy-or”)

- Accounts for noise in visual input

Austerweil & Griffiths (in submission)
Modeling Unitization

\[ x_1 \quad x_2 \quad x_3 \quad x_4 \]

Austerweil & Griffiths (in submission)
Modeling Unitization

\[ X_1 \quad X_2 \quad X_3 \quad X_4 \]

bias \hspace{1cm} learned features

Austerweil & Griffiths (in submission)
Modeling Differentiation

Categorization added to model by adding 70 bits to the end of each image:

- image 1...1 0...0 (category 1)
- image 0...0 1...1 (category 2)
Modeling Differentiation

Categorization added to model by adding 70 bits to the end of each image:
- image 1...1 0...0 (category 1)
- image 0...0 1...1 (category 2)

Objects and Human Learned Features

Austerweil & Griffiths (in submission)
Modeling Differentiation

Categorization added to model by adding 70 bits to the end of each image:

image 1...1 0...0 (category 1)
image 0...0 1...1 (category 2)

Objects and Human Learned Features

Features Inferred By Model

Austerweil & Griffiths (in submission)
Modeling Differentiation

Categorization added to model by adding 70 bits to the end of each image:
- image 1...1 0...0 (category 1)
- image 0...0 1...1 (category 2)

Objects and Human Learned Features

Austerweil & Griffiths (in submission)
Modeling Differentiation

Categorization added to model by adding 70 bits to the end of each image:
- image 1...1 0...0 (category 1)
- image 0...0 1...1 (category 2)

Objects and Human Learned Features

Austerweil & Griffiths (in submission)
Conclusions and Future Directions
Conclusions and Future Directions

• The perceptual system changes its features to find an optimal representation for the objects it observes
Conclusions and Future Directions

• The perceptual system changes its features to find an optimal representation for the objects it observes

• Incorporating (or inferring) Gestalt principles
Conclusions and Future Directions

• The perceptual system changes its features to find an optimal representation for the objects it observes

• Incorporating (or inferring) Gestalt principles

• More principled inclusion of categorization
Conclusions and Future Directions

- The perceptual system changes its features to find an optimal representation for the objects it observes.
- Incorporating (or inferring) Gestalt principles.
- More principled inclusion of categorization.
- Effects on context on the inferred features.
- Two feature representations of macaroni pasta.
Thank you!

We would also like to thank:

• Rob Goldstone
• Stephen Palmer
• Karen Schloss
• The CoCoSci Lab
• The Air Force Office of Scientific Research and the UC Berkeley Chancellor’s Partnership Faculty Fund for $$