Learning hypothesis spaces and dimensions through concept learning

By Joe Austerweil and Tom Griffiths
Department of Psychology, UC Berkeley

http://cocosci.berkeley.edu/
is a gnarble?

gnarble?

gnarble?
\[ g_{ij} \propto e^{-d_{ij}} \]

\[
d_{ij} = \left( \sum_{k=1}^{K} |x_{ik} - x_{jk}|^r \right)^{1/r}
\]

Shepard (1987)
How do you represent a new domain in psychological space?

Shepard (1987)
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
Using the Bayesian generalization model
Experiment - Human hypothesis space learning

Conclusions
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
 Using the Bayesian generalization model
 Experiment - Human hypothesis space learning

Conclusions
A circle with a radius of size 60 mm is a gnarble?
What other sizes of circles are gnarbles?
A circle with a radius of size 60 mm is a gnarble? What other sizes of circles are gnarbles?

\[ h = [a, b] \quad p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{o.w.} \end{cases} \quad P(h|x) \propto P(x|h)P(h) \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a gnarble?
What other sizes of circles are gnarbles?

\[ h = [a, b] \quad p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{o.w.} \end{cases} \quad P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h:y \in h} P(h|x) \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a gnarble? What other sizes of circles are gnarbles?

\[ h = [a, b] \quad p(x|h) = \begin{cases} 
\frac{1}{|h|} & x \in h \\
0 & \text{o.w.}
\end{cases} \]

\[ P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h:y \in h} P(h|x) \quad p(y \in C|x) \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a gnarble? What other sizes of circles are gnarbles?

\[ h = [a, b] \quad p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{otherwise} \end{cases} \quad P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h: y \in h} P(h|x) \]

\[ y = 60 \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a gnarble? What other sizes of circles are gnarbles?

\[ h = [a, b] \quad p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{o.w.} \end{cases} \quad P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h:y \in h} P(h|x) \]

\[ y = 61 \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a *gnarble*? What other sizes of circles are *gnarbles*?

\[ h = [a, b] \]

\[ p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{o.w.} \end{cases} \]

\[ P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h:y \in h} P(h|x) \]

\[ y = 62 \]

Shepard (1987), Tenenbaum & Griffiths (2001)
A circle with a radius of size 60 mm is a *gnarble*?
What other sizes of circles are *gnarbles*?

\[ h = [a, b] \quad p(x|h) = \begin{cases} \frac{1}{|h|} & x \in h \\ 0 & \text{o.w.} \end{cases} \quad P(h|x) \propto P(x|h)P(h) \]

\[ P(y \in C|x) = \sum_{h:y \in h} P(h|x) \]

\[
y = 63
\]
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
Using the Bayesian generalization model
Experiment - Human hypothesis space learning

Conclusions
Similarity and 2-dimensional distance

- Brightness
- Saturation
- Integral ($r = 2$)
- Rotation
- Size
- Separable ($r = 1$)

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

Brightness

Saturation

Integral ($r = 2$)

Rotation

Size

Separable ($r = 1$)

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

Saturation

Brightness

Integral \((r = 2)\)

Rotation

Size

Separable \((r = 1)\)

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

Brightness

Saturation

Rotation

Size

Integral (r = 2)

Separable (r = 1)

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

Saturation

Brightness

Integral (r = 2)

Rotation

Size

Separable (r = 1)

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

Shepard (1958; 1964, 1987)

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]

Brightness

Saturation

Integral \((r = 2)\)

Rotation

Size

Separable \((r = 1)\)
Similarity and 2-dimensional distance

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]

Shepard (1958; 1964; 1987)
Similarity and 2-dimensional distance

Saturation

Brightness

Integral (r = 2)

Rotation

Size

Separable (r = 1)

Integral (r = 2)

Separable (r = 1)

Shepard (1958; 1964, 1987)

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]
Similarity and 2-dimensional distance

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]

Shepard (1958; 1964, 1987)
Similarity and 2-dimensional distance

**Integral (r = 2)**

\[ d_{ij} = \sqrt{\sum_{k=1}^{2} (x_{ik} - x_{jk})^2} \]

**Separable (r = 1)**

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]

Shepard (1958; 1964, 1987)
Similiarity and 2-dimensional distance

Integral (r = 2)

\[ d_{ij} = \sqrt{\sum_{k=1}^{2} (x_{ik} - x_{jk})^2} \]

Shepard (1958; 1964, 1987)

Separable (r = 1)

\[ d_{ij} = \sum_{k=1}^{2} |x_{ik} - x_{jk}| \]
What about 2-D?

Two possible hypothesis spaces:

- \( r = 2 \) (Integral)
- \( r = 1 \) (Separable)

Shepard (1987), Davidenko & Tenenbaum (2001)
How do you know whether a domain is separable or integral?
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
Using the Bayesian generalization model
Experiment - Human hypothesis space learning

Conclusions
Learning hypothesis spaces

Hypothesis spaces \( \mathcal{H}_1, \ldots, \mathcal{H}_M \)

Observe previous concepts \( C = x_1, \ldots, x_n \)

Observe part of new concept \( x_{n+1} \)

How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?
Learning hypothesis spaces

Hypothesis spaces

Observe previous concepts

Observe part of new concept

How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?

\[ P(\mathcal{H}_m | \mathcal{C}, \mathbf{x}_{n+1}) = \frac{P(\mathcal{C}, \mathbf{x}_{n+1} | \mathcal{H}_m) P(\mathcal{H}_m)}{\sum_{i=1}^{M} P(\mathcal{C}, \mathbf{x}_{n+1} | \mathcal{H}_i) P(\mathcal{H}_i)} \]
Learning hypothesis spaces

Hypothesis spaces \( \mathcal{H}_1, \ldots, \mathcal{H}_M \)
Observe previous concepts \( \mathcal{C} = x_1, \ldots, x_n \)
Observe part of new concept \( x_{n+1} \)

How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?

\[
P(\mathcal{H}_m | \mathcal{C}, x_{n+1}) = \frac{P(\mathcal{C}, x_{n+1} | \mathcal{H}_m)P(\mathcal{H}_m)}{\sum_{i=1}^{M} P(\mathcal{C}, x_{n+1} | \mathcal{H}_i)P(\mathcal{H}_i)}
\]

Likelihood under \( \mathcal{H}_m \)
How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?

\[
P(H_m|C, x_{n+1}) = \frac{P(C, x_{n+1}|H_m)P(H_m)}{\sum_{i=1}^{M} P(C, x_{n+1}|H_i)P(H_i)}
\]
Learning hypothesis spaces

Hypothesis spaces \( \mathcal{H}_1, \ldots, \mathcal{H}_M \)
Observe previous concepts \( C = x_1, \ldots, x_n \)
Observe part of new concept \( x_{n+1} \)

How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?

\[
P(\mathcal{H}_m | C, x_{n+1}) = \frac{P(C, x_{n+1} | \mathcal{H}_m)P(\mathcal{H}_m)}{\sum_{i=1}^{M} P(C, x_{n+1} | \mathcal{H}_i)P(\mathcal{H}_i)}
\]

Likelihood under \( \mathcal{H}_m \)

Prior prob. of \( \mathcal{H}_m \)

Normalize over all hypothesis spaces
Learning hypothesis spaces

Hypothesis spaces \( \mathcal{H}_1, \ldots, \mathcal{H}_M \)

Observe previous concepts \( C = x_1, \ldots, x_n \)

Observe part of new concept \( x_{n+1} \)

How should an ideal observer update her belief in hypotheses and generalizations after observing a set of concepts?

\[
P(\mathcal{H}_m|C, x_{n+1}) = \frac{P(C, x_{n+1}|\mathcal{H}_m)P(\mathcal{H}_m)}{\sum_{i=1}^M P(C, x_{n+1}|\mathcal{H}_i)P(\mathcal{H}_i)}
\]

Prior prob. of \( \mathcal{H}_m \)

Likelihood under \( \mathcal{H}_m \)

Normalize over all hypothesis spaces

\[
P(y|C, x_{n+1}) = \sum_{m=1}^M P(y|\mathcal{H}_m, x_{n+1})P(\mathcal{H}_m|C, x_{n+1})
\]
### Integral vs. separable from categorization training

<table>
<thead>
<tr>
<th>Height</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Separable

![Separable diagram](image)

#### Integral

![Integral diagram](image)

Every dot is an object varying in two dimensions.

Each line is a category.
Integral vs. separable from categorization training

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Separable

Integral

Every dot is an object varying in two dimensions.

Each line is a category.
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
Using the Bayesian generalization model
Experiment - Human hypothesis space learning

Conclusions
Experiment - Procedure

Training: For each category, participants classify whether or not every other object is a member of that category.

Are gnarbles gnarble?
Training: For each category, participants classify whether or not every other object is a member of that category.

Test: Given a single exemplar for each category, participants classify every other object.
Experiment - Results

Integral - Separable Model Predictions

Filled = integral > separable
Open = separable > integral
Experiment - Results

Integral - Separable Model Predictions

Filled = integral > separable
Open = separable > integral
Experiment - Results

Integral - Separable Model Predictions

Filled = integral > separable
Open = separable > integral
**Experiment - Results**

- **Integral - Separable Model Predictions**

- **Integral (n=18) - Separable (n=15) Experiment Results**

**Legend:**
- Filled = integral > separable
- Open = separable > integral
Experiment - Results

Integral - Separable Model Predictions

Filled = integral > separable
Open = separable > integral

Integral (n=18) - Separable (n=15)

Difference in width

Difference in height
Experiment - Results

Integral - Separable Model Predictions

Integral (n=18) - Separable (n=15)

Experiment Results

Filled = integral > separable
Open = separable > integral
The Bayesian generalization model

Generalization in two dimensions

Learning hypothesis spaces:
Using the Bayesian generalization model
Experiment - Human hypothesis space learning

Conclusions
Conclusions

Being close in psychological space depends on the distance metric used.

Many stimuli use either city-block or Euclidean distance.

Learning a hypothesis space is akin to choosing a distance metric.

The distance metric may not be a fixed property of the stimuli.

Concept learning affects the distance metric people use.

We demonstrate that integral and separable can be a mutable property of a stimulus.
Acknowledgements

• Nick Chater and three anonymous reviewers
• Tania Lombrozo
• Karen Schloss
• Stephen Palmer
• Rob Goldstone
• RAs: David Belford, Brian Tang, Shubin Li, Ingrid Liu, Julia Ying
• CoCoSci, Concepts and Cognition Lab
• You!
Integral vs. separable from categorization training

\[ P(\mathcal{H}_{\text{sep}} | C) > P(\mathcal{H}_{\text{int}} | C) \]

\[ P(\mathcal{H}_{\text{int}} | C) > P(\mathcal{H}_{\text{sep}} | C) \]