“Second-Order Learning” as a Source of Structure Stabilization in Both Individual Learning and Cultural Evolution

Colin Dawson
The University of Arizona
Outline

• Brief review of some learning mechanisms
• Describe potential sources of constraints on learning
• Explore ways infants might restrict set of generalizations they consider by drawing on prior experience in a domain-dependent way
  – Data from artificial grammar learning
• Speculate wildly about the implications for cultural evolution
A Classic Problem for Learning

• Any data set can be captured by an infinite number of generalizations
  – E.g., “2 4 8 16 32 64”
    • Integer powers of 2?
    • Even numbers?
    • Numbers less than 100?
    • Integers?
    • Black squiggles?
A Role for Statistics

• In the absence of deductive proof, statistical inference may be the best tool to form generalizations
Statistical Learning

• SL has been implicated in a number of learning areas related to language:
  – Segmentation (Saffran and colleagues)
  – Phonetic category learning (Maye & Gerken)
  – Sequential dependencies in an FSG (Gómez & Gerken, 1999)
Statistical Learning

- However, even if requirement for deductive proof is relaxed, problem space is still prohibitively large
• However, even if requirement for deductive proof is relaxed, problem space is still prohibitively large
Statistical Learning

• However, even if requirement for deductive proof is relaxed, problem space is still prohibitively large
Statistical Learning

• However, even if requirement for deductive proof is relaxed, problem space is still prohibitively large
Statistical Learning

• However, even if requirement for deductive proof is relaxed, problem space is still prohibitively large
Statistical Learning

• It gets even worse…

bapikutilofa
A Need for Constraints

• Must be some source of restriction on possible generalizations, representations
Several Possible Classes of Constraints
Several Possible Classes of Constraints

• Perceptual/Representational Constraints
  – Cannot learn what you cannot represent
Several Possible Classes of Constraints

• “Gestalt” principles
  – Meaningful units tend to comprise continuous regions in space and time
Several Possible Classes of Constraints

• “Gestalt” principles
  – Meaningful units tend to comprise continuous regions in space and time
  – Gómez (2002): learners preferentially learn adjacent dependencies -- will only learn non-adjacent dependencies when adjacent ones sufficiently unreliable
Several Possible Classes of Constraints

• Multiple Converging Cues
  – Seems to be widespread in learning
  – Gerken, Wilson, & Lewis (2005)
Several Possible Classes of Constraints

• Prior knowledge about domains
  – Could take the form of rich, innate knowledge structures (e.g. Principles and Parameters)
Several Possible Classes of Constraints

• Prior knowledge about domains
  – Could take the form of rich, innate knowledge structures (e.g. Principles and Parameters)
  – Could be more gradient, but still innate domain-specific biases
Example From “Rule-Learning”

• Marcus, Vijayan, Bandi Rao, & Vishton (1999); Marcus, Fernandes and Johnson (2007)
  – Familiarize 7m infants with several sequences, each of which has a particular abstract pattern (AAB or ABB)
Example From “Rule-Learning”

- Marcus, Vijayan, Bandi Rao, & Vishton (1999); Marcus, Fernandes and Johnson (2007)
  - Familiarize 7m infants with several sequences, each of which has a particular abstract pattern (AAB or ABB)
  - Test on novel sequences, measure looking times to abstractly familiar, abstractly novel sequences
Example From “Rule-Learning”

• Marcus, Vijayan, Bandi Rao, & Vishton (1999); Marcus, Fernandes and Johnson (2007)
  – Familiarize 7m infants with several sequences, each of which has a particular abstract pattern (AAB or ABB)
  – Test on novel sequences, measure looking times to abstractly familiar, abstractly novel sequences
  – Finding is that when elements are syllables, 7-month-olds successfully discriminate, but not when elements are tones, animal sounds
Example From “Rule-Learning”

• Marcus, Vijayan, Bandi Rao, & Vishton (1999); Marcus, Fernandes and Johnson (2007)
  – They argue that speech is privileged somehow, wrt abstract “rule-like” structure
Example From “Rule-Learning”

• Marcus, Vijayan, Bandi Rao, & Vishton (1999); Marcus, Fernandes and Johnson (2007)
  – They argue that speech is privileged somehow, wrt abstract “rule-like” structure

• Alternative (not mutually exclusive) possibility:
  – Infants take advantage of prior experience with each domain to constrain the generalizations they consider
Learning Domain Structure: Language
Learning Domain Structure: Language

- Phonetic tuning (Werker and colleagues)
  - Infants restrict phonetic discrimination to selectively perceive native phonological contrasts
Learning Domain Structure: Language

• Learning “natural” stress rules (Gerken and Bollt)
  - 9m infants learn “stress heavy syllables”, but not “stress syllables beginning with /t/”
Learning Domain Structure: Language

- Learning “natural” stress rules (Gerken and Bollt)
  - 9m infants learn “stress heavy syllables”, but not “stress syllables beginning with /t/”
  - 7.5m infants can learn “stress syllables beginning with /t/”
Learning Domain Structure: Language

• Learning “natural” stress rules (Gerken and Bollt)
  – 9m infants learn “stress heavy syllables”, but not “stress syllables beginning with /t/”
  – 7.5m infants can learn “stress syllables beginning with /t/”
  – Can be attributed to infants’ experience with English, where heaviness, not onset, reliable cue for stress
Learning Domain Structure: Music
Learning Domain Structure: Music

- Reliance on relative over absolute pitch for segmentation (Saffran and colleagues)
  - Intervals between pitches, not overall frequencies, define melodies

+4  +1  -4  +2  -6

A  C#  D  Bb  C  G
Learning Domain Structure: Music

- Learning to attend to culture appropriate tonality/rhythmic structure (Trainor and Trehub, Hannon and Trehub)
Learning Domain Structure: Music

• Learning to attend to culture appropriate tonality/rhythmic structure (Trainor and Trehub, Hannon and Trehub)
  – Infants, not adults, sensitive to melodic alterations that remain within scale (both sensitive when scale violated)
Learning Domain Structure: Music

• Learning to attend to culture appropriate tonality/rhythmic structure (Trainor and Trehub, Hannon and Trehub)
  – Infants, not adults, sensitive to melodic alterations that remain within scale (both sensitive when scale violated)
Learning Domain Structure: Music

• Learning to attend to culture appropriate tonality/rhythmic structure (Trainor and Trehub, Hannon and Trehub)
  – Infants, not adults, sensitive to melodic alterations that remain within scale (both sensitive when scale violated)
Learning Domain Structure: Music

- Learning to attend to culture appropriate tonality/rhythmic structure (Trainor and Trehub, Hannon and Trehub)
  - Infants, not adults, sensitive to melodic alterations that remain within scale (both sensitive when scale violated)
  - Infants, not (North American) adults, sensitive to rhythmic changes that preserve small integer ratios
Learning Domain Structure

- Perhaps infants learn to attend to relationships characteristic of particular domains
- What merits attention in one domain may not in another
Two Predictions

• To the extent that domain-specificity is learned:
Two Predictions

• To the extent that domain-specificity is learned:
  – Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
Two Predictions

- To the extent that domain-specificity is learned:
  - Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  - Domain-naïve learners’ abilities will be better predicted by more general factors (e.g., stimulus complexity, memory demands)
Experiment 1
Experiment 1

• Familiarize 18 4-month-old infants with phrases consisting of three chords
  – Half of infants familiarized with AAB, half with ABA
  – Four different “A” chords, four different “B” chords, occur in all combinations
    • Both “A”s and “B”s evenly split between major and minor triads
  – Balanced between pitch contours (rise vs. fall)
Experiment 1

• Test on novel phrases, containing two new “A” chords, two new “B” chords, in all combinations
  – Again split between major and minor triads
  – Again balanced between rise vs. fall
Experiment 1

- 4-month-olds look longer during test trials from novel grammar
Experiment 2

- Want to make sure the effect is not driven by stimulus idiosyncracies
- Test 7.5-month-olds with same stimuli
Experiment 2

- 7.5-month-olds do not succeed, replicating MFJ ‘07
Experiment 3

• Test 4-month-olds on single tones, to provide more direct comparison to MFJ ‘07
Experiment 3

- 4-month-olds looking times to single-tone stimuli
Experiments 1-3

**Mean Looking Times by Experiment and Consistency**

- **4m-Chords**
- **7.5m-Chords**
- **4m-Tones**

<table>
<thead>
<tr>
<th>Consistency</th>
<th>Mean LT in Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>*</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Two Predictions

• To the extent that domain-specificity is learned:
  – Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Domain-naïve learners’ abilities will be better predicted by more general factors (e.g., stimulus complexity, memory demands)
Two Predictions

• To the extent that domain-specificity is learned:
  – Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Domain-naïve learners’ abilities will be better predicted by more general factors (e.g., stimulus complexity, memory demands)
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
  – Should succeed at domain-typical generalizations
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
  – Should succeed at domain-typical generalizations
  – But what is actually typical and atypical of music?
Musical Corpus Analysis

• Prediction: Relationship of phrase-final chord to key reliable cue; serial identity relationships unreliable

\[ \text{C E E G D D C} \]

\[ 1 \ 3 \ 3 \ 5 \ 2 \ 2 \ 1 \]

\[ \text{diff \ same \ diff \ diff \ same \ diff} \]
Musical Corpus Analysis

• Prediction: Relationship of phrase-final chord to key reliable cue; serial identity relationships unreliable
Musical Corpus Analysis

- 22 most popular (> 100K Google hits of title) songs selected from book of popular children’s songs (Winn and Miller, 1966)
Musical Corpus Analysis

• 22 most popular (> 100K Google hits of title) songs selected from book of popular children’s songs (Winn and Miller, 1966)
• Melody lines parsed into phrases each two measures long
Musical Corpus Analysis

• 22 most popular (> 100K Google hits of title) songs selected from book of popular children’s songs (Winn and Miller, 1966)
• Melody lines parsed into phrases each two measures long
• Each phrase coded for repetition pattern of last three notes (AAA, AAB, ABA, ABB, ABC), as well as final chord (printed above last note)
## Musical Corpus Analysis

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>8 (5.8%)</td>
</tr>
<tr>
<td>AAB</td>
<td>23 (16.7%)</td>
</tr>
<tr>
<td>ABA</td>
<td>12 (8.7%)</td>
</tr>
<tr>
<td>ABB</td>
<td>17 (12.3%)</td>
</tr>
<tr>
<td><strong>Total w/ rep</strong></td>
<td><strong>60 (43.5%)</strong></td>
</tr>
<tr>
<td><strong>no reps (ABC)</strong></td>
<td><strong>78 (56.5%)</strong></td>
</tr>
<tr>
<td>ends in I</td>
<td>90 (65.2%)</td>
</tr>
<tr>
<td>ends in V</td>
<td>38 (27.5%)</td>
</tr>
<tr>
<td><strong>ends in I or V</strong></td>
<td><strong>128 (92.7%)</strong></td>
</tr>
<tr>
<td>other</td>
<td>10 (7.2%)</td>
</tr>
</tbody>
</table>
Statistical Tests

• Chi-square goodness of fit tests:
  – Repetition (collapsing across all types) occurs statistically at chance (p > 0.25)
  – Phrases ending in I and ending in V each occur more often than chance (p < 10^{-64} and p < 10^{-5} respectively)
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
• Can interpret results of Exp. 2, MFJ ‘07 as due to atypicality of serial identity relations in music
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
    • Can interpret results of Exp. 2, MFJ ‘07 as due to atypicality of serial identity relations in music
  – Should succeed at domain-typical generalizations
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
    • Can interpret results of Exp. 2, MFJ ‘07 as due to atypicality of serial identity relations in music
  – Should succeed at domain-typical generalizations
    • Supported for language by Marcus, et al. (1999)
Two Predictions

• Domain-typicality of a generalization will predict domain-savvy learners’ abilities to learn that structure
  – Should fail at domain-atypical generalizations
    • Can interpret results of Exp. 2, MFJ ‘07 as due to atypicality of serial identity relations in music
  – Should succeed at domain-typical generalizations
    • Supported for language by Marcus, et al. (1999)
    • But what about domain-typical generalization for music?
Experiment 4

- Familiarize 7.5-month-olds with melodies that are constant with respect to scale-degree of last note
Experiment 4

• Familiarize 7.5-month-olds with melodies that are constant with respect to scale degree of last note
  – Half familiarized with melodies ending on I (“do”), half ending on V (“sol”)
  – Eight different carrier melodies, each in a different key
  – Melodies prepended with I-V-I chord sequence to establish key
Experiment 4

• In “ends-in-I” grammar, I note appended to each carrier melody, V note for “ends-in-V” grammar

• Four new carrier melodies composed for test phase, in new keys
  – Same carrier melodies for “e1” trials and “e5” trials
Experiment 4

- 7.5-month-olds displayed significant preference for novel end chord
Interim Summary

• Younger infants appear to learn AAB vs. ABA in chords and tones
  – Inconsistent with idea that speech is privileged from the beginning
Interim Summary

• Younger infants appear to learn AAB vs. ABA in chords and tones
  – Inconsistent with idea that speech is privileged from the beginning

• Older infants’ ability to learn relational generalizations in music appears to be related to reliability of type of relation in input
  – Similar to shape vs. material bias for objects vs. substances (Smith and colleagues)
Interim Conclusions

• Perhaps domain-specificity emerges as learners notice that low level features predict high-level structure
  – Represents “middle ground” between innate domain-specificity and completely domain-general learning mechanisms
Interim Conclusions

• Perhaps domain-specificity emerges as learners notice that low level features predict high-level structure
  – Represents “middle ground” between innate domain-specificity and completely domain-general learning mechanisms
• Might operationalize notion of “domain” to mean sets of environments across which same kinds of cues produce adaptive generalizations
Mental model of environment
Mental model of environment
Acknowledgements

• LouAnn Gerken
• Juliet Minton
• Diana Moreno
• NSF GRFP to CD
• NIH R01 HD042170 to LAG