CBCD turns 21: What have we learned about the mechanisms of learning and development?

Richard N. Aslin

Haskins Laboratories, Yale University, University of Connecticut
Linkages to my past

• Gyorgy Gergely: visiting professor at Rochester (1989-90)
• Mark Johnson and Annette K-S: McDonnell/PEW task force (1998-99)
• Denis Mareschal and Leslie Tucker
• 2006-07 CBCD sabbatical
• Natasha Kirkham: student of Scott Johnson (postdoc at Rochester)
• Rachel Wu: postdoc at Rochester
Organization of Behavior

by

D. O. Hebb

Stimulus and response — and what occurs in the brain in the interval between them

$3.95

(1949)
The Relation of Early to Later Learning: “It is of course a truism that learning is often influenced by earlier learning. Innumerable experiments have shown such a 'transfer of training'. Learning A may be speeded up, hindered, or qualitatively changed by having learned B before. . . . If the learning we know and can study, in the mature animal, is heavily loaded with transfer effects, what are the properties of the original learning from which those effects came? How can it be possible even to consider making a theory of learning in general from the data of maturity only? There must be a serious risk that what seems to be learning is really half transfer.” (pp.109-110)
Learning vs. Development

• Is development merely the historical outcome of learning at the age when developmental status is assessed?

• Is learning cumulative (i.e., bigger, better, faster) or does it lead to qualitative change, and by what mechanism?
  - Adding or deleting an underlying structure or process
  - Unmasking an existing structure or process (e.g., via noise reduction)

• What is special about development that is not captured by learning?

• Early learning shapes later learning by facilitating or constraining it
Classic example: Imprinting

- Konrad Lorenz: Nobel Prize in 1973
Sensitive period for perception in humans

3-D vision requires proper eye alignment

Strabismus
Early experience (eye-alignment) matters

- Banks, Aslin & Letson (*Science*, 1975)
Sensitive period for language?

- Henry Kissinger, b. Germany 1923
- Moved to U.S. in 1938
- Speaking English for 81 years
- Johnson & Newport (1989): speech errors made by people who learned English at different ages

The later you begin learning a second language, the more errors you make.
The curse of developmental plasticity

- If early experience is atypical and plasticity declines with age, then learning mechanisms may not be able to recover from early errors.
- Too much plasticity $\rightarrow$ only most recent input matters.
- Too little plasticity $\rightarrow$ slow time-course of adapting to changing input.
- **Trade-off:** *explore* the environment to gather new information vs. *exploit* what has already been learned to become an efficient user of that information.
- **Dilemma:** the bias to exploit works well in a stationary world, whereas the bias to explore works well in a highly volatile world.
- Complex generative models work best when stationarity/volatility is in balance $\rightarrow$ time-course of learning matches minimal epochs of stationarity.
Outline

1. What are the dominant historical trends in infancy research?
2. Highlight the robustness and flexibility of learning in infants
3. What have measures of brain activity revealed about the mechanisms of development?
4. Do naïve learners integrate prior information with current input?
1. What are the dominant historical trends in infancy research?
Scientists discover the “competent infant”

• In the 1950’s, most major textbooks in Ophthalmology stated that newborns were “blind”
• Most major textbooks in Audiology stated that newborns were ”deaf”
• Over the subsequent 30 years, infants were found to have remarkable, although limited, perceptual abilities
• Over the past 30 years, it was confirmed that infants have amazing abilities to learn their native language, the properties of objects, and how other people think and reason in social settings
• Are “modern” babies just smarter than their ancestors?
What drove this increase in knowledge about infants?

What’s in a look?

Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, USA

Abstract

The most common behavioral technique used to study infant perception, cognition, language, and social development is some variant of looking time. Since its inception as a reliable method in the late 1950s, a tremendous increase in knowledge about infant competencies has been gained by inferences made from measures of looking time. Here we examine the logic, utility, and future prospects for further gains in our understanding of infant cognition from the use of looking time measures.
Development is rapid and transformational

William Kessen, Marshall Haith & Philip Salapatek

“Whether one sees the newborn child as neurologically insufficient (Flechsig, 1920), cognitively confused (James, 1890), narcissistic (Freud, 1905), solipsistic (Piaget, 1927), or merely ugly (Hall, 1891), the distance between the new child and the walking, talking, socially discriminating, and perceptive person whom we see hardly 500 days later is awesome.”
1965 – 1985: what are infants’ capacities?
1985 – 2005: how are capacities utilized?
2005 – 2025: how do infants form causal (generative) models and apply them in the natural environment?

Discovery of latent variables allows a learner to transfer knowledge (generalize) to novel contexts.
Learning by action and by observation

85%

Why 15% failure?
Jammed mechanism
Late restock delivery
Power failure
Pounding/shaking
Learning is fundamentally an inference problem

- Events are probabilistic
- Associations are only one source of information and are often spurious
- Causes are often hidden
- Ambiguity is ubiquitous
- 1 3 5 7 __
- \( \text{sum2} + 1, \text{sum2} - 1, \text{etc.} = 13 \)
How do infants form causal (generative) *models*?

Alison Gopnik: blicket detector

Other examples of infants’ flexibility and transfer of learning
2. Highlight the robustness and flexibility of learning in infants
White & Aslin (2011)

Known-word object labeling pre-test phase

18-month-olds

Look longer at match

Even if match was just mispronounced

Same for Unlabeled

bettle??

block
bleck
A-not-B task
10-month-olds

Ostensive-communicative  Noncommunicative
Xu & Garcia (2008)

If person drawing samples has a strong preference for white balls (i.e., less likely outcome), then looking pattern is reversed

Xu & Denison (2009)
Téglás, Girotto, Gonzalez, & Bonatti (2007)
Choosing the “right” inference

Occam’s Razor: Among competing hypotheses, the one with the fewest assumptions should be selected.  
Wikipedia

This does not ensure that the simplest explanation is the correct explanation, but if you have no other basis for preferring explanation #1 over explanation #2, go with simplicity.

Caveat: Simple could be wrong – Michael Maratsos
Model-free  

Model-based

Doll, Simon & Daw (2012)
Team Smart vs. Team Dumb

• Model-based (e.g., Rule learning) or Model-free (e.g., Statistical learning)
• How does Smart emerge from Dumb?
• Two separate mechanisms or a continuum?
• What patterns of input license “abstraction”?  
  • Specialized triggers (innate or maturational)  
  • General principles (domains, modalities, species)
Marcus et al. (1999): rule learning

- 3-element *strings*
- Pauses between strings enable encoding of position
- Syllable ‘identity’ used to define category
Gerken (2006): Context-specific generalization (broad vs. narrow)

Syllable B

\[ \begin{array}{c|c|c|c}
\text{Di} & \text{Je} & \text{Li} & \text{We} \\
\end{array} \]

Syllable A

\[ \begin{array}{c}
\text{Le} & \text{Wi} & \text{Ji} & \text{De} \\
\end{array} \]

\[ \begin{array}{c|c|c|c|c|c|c}
\text{LeLeDi} & \text{LeLeJe} & \text{LeLeLi} & \text{LeLeWe} \\
\text{WiWiDi} & \text{WiWiJe} & \text{WiWiLi} & \text{WiWiWe} \\
\text{JiJiDi} & \text{JiJiJe} & \text{JiJiLi} & \text{JiJiWe} \\
\text{DeDeDi} & \text{DeDeJe} & \text{DeDeLi} & \text{DeDeWe} \\
\end{array} \]
6-year-olds can learn word-category rules
Schuler, Reeder, Kissinger, Aslin & Newport (2017)

- Listen to a set of 3-word sentences
- Words are assigned to 3 categories (3 words per category)

```
<table>
<thead>
<tr>
<th>A</th>
<th>X</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>X₁</td>
<td>B₁</td>
</tr>
<tr>
<td>A₂</td>
<td>X₂</td>
<td>B₂</td>
</tr>
<tr>
<td>A₃</td>
<td>X₃</td>
<td>B₃</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
```

“klidum bleggin glim.”    “mawg fluggit zub.”
Space Alien language
Test on novel (withheld) sentences
6-year-olds can learn rules and exceptions
6-month-olds looked significantly more at named target images when the competitor images were semantically unrelated (e.g., milk and foot) than when they were related (e.g., milk and juice)

Bergelson & Aslin (2017)
David Pisoni: “One replication is worth 1000 t-tests”

SeedLings Project: Duke University

Word recognition

Number of times that word is spoken while infant is looking at object

\[ r = .39, p = .013 \]
3. What have measures of brain activity revealed about the mechanisms of development?
What is the “value added” of brain over behavior?

• Development of behavior typically lags brain development → earlier diagnosis and more sensitive assessment of treatment effectiveness

• Behavioral development exhibits qualitative change → implies a fundamental change in brain structure or function

✓ At 9 months, begin to search for hidden object
✓ At 18 months, begin to produce 2-word sentences
What are the brain recording techniques?

- EEG
  [Image of EEG setup]
  Courtesy of Patricia Kuhl

- MEG
  [Image of MEG setup]

- NIRS
  [Image of NIRS setup]

- MRI
  [Image of MRI machine]
  Courtesy of John Richards
Diffuse optical imaging of relative changes in oxy- and deoxy-hemoglobin concentrations
Pros and Cons

• No acoustic noise from pulse sequences
• Tolerates considerable head motion (emitters and detectors fixed to a cap)
• Allows upright posture and a variety of natural responses
• Limited to cortical surface (sub-cortical and ventral cortical areas not accessible)
• Coarse spatial resolution (~ 1 cm²; not 3D voxel)
• Deoxy-hemoglobin (BOLD) signal is weak
• Surface vasculature creates non-cortical “systemic noise”
The Predictive Coding Perspective

• Sensory signal that is transformed in a feed-forward hierarchy
• Biased interpretation (via priors) of sensory signal → prediction
• Discrepancy between sensory signal and prediction → prediction error

• How close to sensory signal does top-down prediction propagate?
Benefits of Predictive Coding

• Time-critical events (e.g., speech, reading) benefit from a reduction of alternatives during on-line processing

• Motor planning and execution requires learning a Forward-Model to compensate for time delays and kinematic constraints

• Updating of a generative model requires efficient brain pathways that can be flexibly deployed depending on context, particularly in a non-stationary (volatile) environment
Stimulus Omission Paradigm

• Hughes et al. (2001)
• ECoG from presurgical epilepsy patients using cortical electrode arrays
• 2-tone pairing with occasional 1-tone (unexpected omission) test trials
Emberson, Richards & Aslin (2015)

Occipital cortex
Temporal cortex
Learning Phase: Auditory-Visual Pairing

18 trials of A + V pairings
Test Phase: Expected Pairs and Visual Omissions

- 80% A + V+
  - Honk!

- 20% A + V -
  - Honk!
  - Unexpected omission
Results

![Oxygenated Hemoglobin (mM mm)]

- **A+V+**
  - Temporal
  - Occipital

- **A+V-**
  - Temporal
  - Occipital

**ROIs**
- Temporal
- Occipital
Control: Auditory and Visual Stimuli never Paired

Honk!
Overall Results

V+ is Expected  Unexpected

Oxygenated Hemoglobin (mM mm)

A+V+  A+V-  A+ control  V+ control

ROIs
- Temporal
- Occipital
Emberson et al. (2017)

- At-risk infants: extremely premature (27-33 weeks gestation)
- Tested at 6 month corrected age on visual omission task
4. Do naïve learners integrate prior information with current input?
Maye, Werker & Gerken (2002)

Looking Time (sec.)

<table>
<thead>
<tr>
<th>Training Distribution</th>
<th>Unimodal</th>
<th>Bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonAlternating</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The graph shows the looking time (in seconds) for different training distributions, comparing alternating and nonalternating conditions. The data indicates a higher looking time in the bimodal condition compared to the unimodal condition, especially for the nonalternating case.
Teinonen, Aslin, Alku & Csibra (2008)
Two visual-articulatory gestures override unimodal distribution

\[ p = .02 \]
What do infants do with prior information?

- Learning more/better/faster
- Explore to learn or verify hypotheses

Stahl & Feigenson (2015)
Sim & Xu (2017)

13-month-olds
Sim & Xu (2017)

13-month-olds
Why do children do with prior information?

- Limited computational mechanism: fail to update priors
- Domain-general constraints: show adult-like integration if simple task
- Bejjanki, Murphy & Aslin (2019)
Weights for sensory and prior information

**Adults**

- Double prior

**6-8 year olds**

- Double prior
- Single prior

Block of 400 trials
4-year-olds fail to update: stationarity bias

• Starling, Reeder & Aslin (2018)
- Stationarity bias (simplicity)
- Reduces memory demand
- Explain away volatility as noise
Compositional bottleneck: WM

• Piantadosi, Palmeri & Aslin (2018)
The Future

• Machine-learning techniques and classification-learning paradigms
  EEG: Bayet, Zinszer, Pruitt, Aslin & Wu (2018)
• Big data: Many Babies ‘X’ Project
• Connectomics (whole-brain functional connectivity during rest and movie watching): Sanchez-Alonso, Rosenberg & Aslin (revision under review)
• Portable MEG

Hill, .... & Brookes, 2019
Nature Communications
Summary

• Pay attention to history (search literature from the past, not just from the present)
• Development = constrained learning
• fNIRS reveals top-down predictive architecture in infants
• Efficient and flexible learning is a balance of: simplicity, domain-general cognitive limitations, and task-related pressure to generalize beyond the input
• See you in 2029!!
Team of Collaborators

- Elissa Newport, Georgetown
- Lauren Emberson, Princeton
- Ben Zinszer, University of Delaware
- Vik Rao Bejjanki, Hamilton College
- Laurie Bayet, American University
- Sara Sanchez-Alonso, Haskins
- Katherine White, Waterloo
- Patty Reeder, Gustavus Adolphus
- Katie Schuler, Penn

Funding Sources
- NIH
- Gates Foundation

- Celeste Kidd, UC Berkeley
- Steve Piantadosi, UC Berkeley
- Elika Bergelson, Duke
- Sarah Starling, DeSales Univ.