Experiments in Developmental AI: Learning Action Schemas and Language in Jean's Room

Carole Beal, Paul Cohen, Yu-Han Chang
Wes Kerr, Daniel Hewlett, Shane Hoversten
Learning and Development Center
Information Sciences Institute
University of Southern California
How do young children learn?
Piaget’s view

- Infants exercise innate schemas, older children exercise learned schemas; all learn by doing
- "Usual" instances are assimilated into schemas, unusual cases force accommodation or change of schemas, common sequences become abstracted gists.
- Exercising schemas is rewarding
- Physical maturation and social interaction scaffold the development of new schemas
- Knowledge accumulates around schemas
Outline

• What Jean learns
• What are schemas?
• How are they learned?
• Jean's Room (TICA v0.0)
  – Task learning
  – Language learning
• Phasing the project for five years
• Which architecture?
What Jean learns

ISIS Jean – small unit tactical sim

Toddler Jean – sim room full of objects

The great accomplishments of toddlerhood include learning to do things in the physical and social worlds and learning language

Learns tactical procedures (gists) by experimentation and (soon) by NL interactions with teacher

Accelerates learning due to transfer

Learns word meanings through NL interactions with teacher

Will learn gists by experimentation and NL interactions with teacher
What Jean learns

ISIS Jean – small unit tactical sim

Toddler Jean – sim room full of objects

The great accomplishments of toddlerhood include learning to do things in the physical and social worlds and learning language.

Example: procedure for sneaking up on an enemy unobserved by crawling instead of running

Example: What "go to the green cone" and "move it to the yellow block" mean
Knowledge Foundations
Image Schemas

- Need something barely more abstract than the raw perceptions of a situated agent, yet sufficient for learning abstract knowledge

- Cognitive linguists and other cognitive scientists suggest *image schemas*

<table>
<thead>
<tr>
<th>Space</th>
<th>up-down, front-back, near-far, center-periphery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>containment, in-out, full-empty, content, surface</td>
</tr>
<tr>
<td>Force</td>
<td>balance, counterforce, compulsion, restraint, enablement, blockage, diversion, attraction</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>merging, collection, splitting, part-whole, mass-count</td>
</tr>
<tr>
<td>Identity</td>
<td>matching, superposition</td>
</tr>
<tr>
<td>Existence</td>
<td>object, process, removal, bounded space</td>
</tr>
</tbody>
</table>

- Intriguing but too vague
- Stronger on spatial configurations (statics) than dynamics
- We are developing an Image Schema Language for knowledge about statics and dynamics
Schemas for learning gists and language

Object schemas
Currently just bags of attributes for probabilistic models for nouns and adjectives

Spatial schemas (aka Static Image Schemas)
3D spatial model providing Regier-Oates-like probabilistic semantics for prepositions

Dynamic/Action schemas
Maps for controllers of state variables vs. time, regions for norms and goals

- move-to
- Failure region
- Goal region
Jean learns features of scenes that predict which trajectories will happen (e.g., the cat is animate) and "splits" maps so all the trajectories on a map are roughly the same.

Said differently, minimize the entropy of state variable transitions within a map.
Learning Action Schemas
Experimental State Splitting Algorithm

\[ \vec{V}_t = x_{1,t}, \ldots, x_{n,t}, x_{n+1} = \text{animate} \]

\[ \vec{V}_t = x_{1,t}, \ldots, x_{n,t}, x_{n+1} = \text{inanimate} \]
Learning Gists

Old way: States are tests containing predicates, arcs are actions, actions aren't interruptible, time isn't explicit, tests occur at beginning or end of actions.

New way: States are maps associated with action schemas, time is explicit, arcs are transitions between actions (i.e., changing what you are doing) predicates are associated with regions, regions have associated probabilities.

In both approaches, gists are sequences of actions that occur with high probability.
Jean's Room: The virtual two-year-old
Dave Gunning's list

• Learning to assemble or manipulate objects by observing a mentor/instructor perform the task;
• Learning new concepts including language, by observation in a simplified world or being told by the mentor;
• Finding a target object through exploration and search of the simplified world;
• Tasks relating to theory of mind;
• Generally, learning any task that might be communicated to BICA by the instructor using the language of a two-year old, about this simplified world.
Learning Language

- Jean learns nouns, verbs, prepositions and adjectives referring to things and relations in Jean's room.
- Jean exploits parse trees, previously-learned words, and knowledge about her room to learn new words.
- Jean learns the meanings of multiple words, simultaneously.

You: Go to the red cube
Jean: What's the red cube
You: <point to the red cube>
You: Go to the blue cube
Jean: What's the blue cube
You: <point to the blue cube>

At this point Jean doesn't know "cube" is shape and "red" and "blue" are colors.
Nouns and adjectives

You: Go behind the red cylinder

Jean doesn't ask when she's relatively certain

You: No

Jean goes to the other thing that she thinks might be a red cylinder

You: Yes
You: Go behind the red cube
Jean: Is this the red cube
You: Yes
Jean: Am I behind it?
You: No
Jean: Am I behind it?
You: No
Jean: Am I behind it?
You: Yes
Learning verbs

You: Go to the cone
Pick it up

Jean: What's the it?

You: <point to cone>

You: Go to the blue cube
Put it down

You: Pick up the blue cube

You: Go to the red cylinder
Put it down

You: Walk to the green cone

Anything that can follow directions like this ought to be able to learn simple procedures from a teacher
Phasing for five years

- **Year 1**: Small comprehension vocabulary (500 words), very small production vocabulary and grammar (20 words), learn simple procedures by play and instruction. Major ontological distinctions in semantic memory. Shared attention.

- **Year 2**: Comprehension vocabulary: 1000 words; production vocabulary and grammar: 200 words; link language and concepts. Learn simple games by instruction (e.g., egg hunt); learn by observing other Jeans; self-goal-directed activities. Dawning of episodic memory.

- **Year 3**: Dawning of symbolic function (e.g., symbolic play, drawing). Occasional abstract language, referring to objects in other rooms and mental states. Self-awareness & emotions related to behavior.

- **Years 4-5**: Comprehension vocabulary 10K words; near-adult production competence. Verbal event description. Theory of mind. Equated for familiarity - adult-like semantic memory and gist extraction.
• We are unaffiliated: thrust D
• Our interest is developmentally plausible learning based on static and dynamic image schemas
• Lots of parts in need of architecture

Our goal: an architecture for learning and development in Jean's Room

Simultaneous, ongoing, opportunistic development of language and activities in a physical and social environment
Who are Jean's parents and playmates?

Jean's Room

Help your Jean
- play with friends
- learn new games
- compete in games
- chat with you
- make new toys
- travel

Can you teach me to find the hidden treasure??