Possible Minds

MAS S66
New Destinations in Artificial Intelligence
Goals and Directions for Future Research

joscha@mit.edu
Overview

• Creating Meaning (see: Sloman, Towards a Gibsonian Model of Vision)

• Criteria for understanding Minds (see Sloman, Exploring the Space of Possible Minds)

• Basic functional organization of a human-like mind
Questions

• How does vision work?
• What are quintessential functions of the mind?
• Which ones are specific for the human mind?
• Which ones could be different?
• What are the basic modules/structures we need to look at?
Perception and Representation

• How are sensory data related to conceptual structures?
• How can conceptual structures be represented?

→ We do not see what we are seeing!
What is perception?

This text is upside down.
This txet is upisde donw
This txet is upisde donw
Thatcher Illusion
Thatcher Illusion
Thatcher Illusion
The Symbolic Space

Reality/Body

Symbolic Space

Imagination/Mind
The Symbolic Space

Symbolic Space

Imagination/Mind
The Symbolic Space

Symbolic Space
The Symbolic Space

Reality?

Symbolic Space
The Symbolic Space

Symbolischer Raum

Realität/Körper

Imagination/Geist

FutureAI

10/20/15
How does perception work?

Magritte: The Human Condition II
How to recognize a tree?

130,000 Polygons
fractal structures
large variance in shape and color with viewing angle

Computergrafik ©XFrog
How to recognize a tree?
How to recognize a tree?

Abstract schemas

- foliage
- leaves
- stem
- connected to ground

FutureAI
Schematic hierarchies

Image recognition tells us something about perceptual mechanisms
How do we perceive reality?

Jim Avignon: ““All my friends””
How do we perceive reality?
How do we perceive reality?
How do we perceive reality?

face schema
How do we perceive reality?
How do we perceive reality?
How do we perceive reality?
How do we perceive reality?

face schema
How do we perceive reality?
The symbolic space

Sensors

Symbolic Space

Imagination/Representation
The symbolic space

Sensors

Symbolic Space

Imagination/ Representation
Perception

→ Perception is not simply constructed from receptor input (bottom up)

→ Verification of hypothesis via sensors

→ No difference between perceptual and imaginative datastructures
Modal vs. amodal representation (Barsalou 99)

Perceptual States

Arbitrary ‘amodal’ Symbols

Transduction

Reference

Memory

Language

Thought

(Chair = c₁)
(Back = b₁)
(Seat = s₁)
(Legs = l₁)

Feature lists,
Semantic networks, Frames, Schemata
Predicate Calculus Sentences

Perceptual States

Analogue ‘modal’ Symbols

Extraction

Reference

Memory

Language

Thought

Images,
Image Schemas,
Perceptual Symbols

Neural Activation
(Conscious Experience)

Neural Activation
(Conscious Experience)
Methods should focus on components and performances necessary for intelligence:

- **Whole, testable architectures**
- **Universal Representations:**
  Grounded neuro-symbolic representations (integrate both symbolic and distributed aspects)
- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...
- **Universal Motivation:**
  Polythematic, adaptive goal identification
- **Emotion and affect**
Components for Cognitive AI

- Universal mental representations
  (compositional + distributed \(\rightarrow\) neurosymbolic)
Components for Cognitive AI

- (Semi-) General problem solving: Operations over these representations

(neural learning, categorization, planning, reflection, consolidation, ...)
Components for Cognitive AI

- Perceptual grounding
Components for Cognitive AI

- Perceptual grounding and action
Components for Cognitive AI

- Perceptual grounding and action
Components for Cognitive AI

- Model of current situation, and protocol of past situations
Components for Cognitive AI

• Model of self

Perception

Protocol memory

Self model

Situation model

Op_1
Op_2
Op_3
Op_4
... Op_n

Action
Components for Cognitive AI

• Abstractions of objects, episodes and types
Components for Cognitive AI

- Anticipation of future developments

```
   Perception
     ↓
   Self model
     ↓
Declarative memory
     ↓
Procedural memory
     ↓

Op1  Op2  Op3  Op4  ...  Op_n
  ↓   ↓   ↓   ↓    ↓
Frame Plans Expect.

Action
```

Expectations
Components for Cognitive AI

- Action selection and executive control

Diagram:

- Longterm memory
- World model
- Mental stage
- Perception
- Action
Components for Cognitive AI

- Action selection and executive control

![Diagram showing the components of cognitive AI]

- Perception
- Action selection
- Action

- Longterm memory
- World model
- Mental stage
Components for Cognitive AI

- Universal motivation: autonomous identification of goals
Components for Cognitive AI

- Emotional modulation and affect

**Diagram:**

- Memory
  - Action selection
    - Action
    - Motivational system: Motive Selection and Decision making
    - Selection threshold
    - Urges/drives
  - Perception
    - Resolution
    - Securing rate
    - Arousal
“Classical Cognitive Architectures” tend to focus on cognition as an isolated problem solving capability.
Possible Minds (Eliezer Yudkowsky)

Biological Minds

Human Minds
Possible Minds

All Possible Minds

Biological Minds

Human Minds
Possible Minds

All Optimization Processes

All Possible Minds

Biological Minds

Human Minds
Possible Minds (not to scale)

All Optimization Processes

All Possible Minds

Biological Minds

Human Minds

Biological Evolution
Kinds of Minds (J. Storrs Hall)

- **Hypo-human**: infrahuman, less-than-human capacity.
- **Diahuman**: human-level capacities in some areas, but still not a general intelligence.
- **Parahuman**: similar but not identical to humans, as for example, augmented humans.
- **Allohuman**: as capable as humans, but in different areas.
- **Epihuman**: slightly beyond the human level.
- **Hyperhuman**: much more powerful than human, superintelligent.
Space of Possible Minds (Aaron Sloman)

- Quantitative VS Structural;
- Continuous VS Discrete;
- Complexity of stored instructions;
- Serial VS Parallel;
- Distributed VS Fundamentally Parallel
Space of Possible Minds (Aaron Sloman)

- Connected to External Environment VS Not Connected;
- Moving VS Stationary;
- Capable of modeling others VS Not capable;
- Capable of logical inference VS Not Capable;
- Fixed VS Re-programmable;
- Goal consistency VS Goal Selection;
- Meta-Motives VS Motives;
- Able to delay goals VS Immediate goal following;
- Static Plans VS Dynamic Plans;
- Self-aware VS Not Self-Aware.
Generality of Human Minds

- ability to cope with varied objects in a domain
- ability to cope with a variety of domains of objects
- ability to perform a variety of tasks in relation to any object

- Self-improvement
- Graceful degradation
Specific Abilities of Human Minds

- Inference, reasoning under uncertainty, reasoning with nonlogical representations (maps, diagrams, models)
- Hypothetical questions („What would happen if . . .?“) for plans, predictions, generalizations
- Insight and understanding rather than brute force
- Communication and co-operation with other intelligent agents
- Coping with inconsistent 'motivators', e.g. goals, tastes, preferences, ethical principles, constraints, etc.
- Self-reflection, ability to explain one's actions
- Generate, or appreciate, aesthetic objects
- Experience bodily sensations.
- Enjoy or dislike experiences, to be amused, angry, excited, irritated, hopeful, disgusted, etc.
<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/21</td>
<td>Possibilities for artificial minds</td>
</tr>
<tr>
<td>9/28</td>
<td>Agents within agents. The Society of Mind (Minsky)</td>
</tr>
<tr>
<td>10/05</td>
<td>Vision as inverse rendering? (Poggio?)</td>
</tr>
<tr>
<td>10/12</td>
<td>Columbus Day (no session)</td>
</tr>
<tr>
<td>10/19</td>
<td>AI as engineering or AI as a science. The Norvig/Chomsky debate</td>
</tr>
<tr>
<td>10/26</td>
<td>The Neocognitron and Deep Learning</td>
</tr>
<tr>
<td>11/2</td>
<td>Universal intelligence. From Solomonoff induction to AIXI</td>
</tr>
<tr>
<td>11/9</td>
<td>AI and Neuroscience</td>
</tr>
<tr>
<td>11/23</td>
<td>Affect and Motivation</td>
</tr>
<tr>
<td>11/30</td>
<td>Measuring the Progress of AI. Benchmarks</td>
</tr>
<tr>
<td>12/7</td>
<td>Closing Discussion. Can we sketch a Map of Future AI research?</td>
</tr>
</tbody>
</table>

*Sessions: see futureai.media.mit.edu*
Components for Cognitive AI

Memory

Resolution

Perception

Action selection

Action

Securing rate

Arousal

Motivational system: Motive Selection and Decision making

Urges/drives

Selection threshold
Layers of Cognition

Reflective

Deliberative

Reactive
# Columns of Cognition

<table>
<thead>
<tr>
<th>Perception</th>
<th>Cognitive Processing</th>
<th>Action</th>
</tr>
</thead>
</table>
### Cognitive Grid

<table>
<thead>
<tr>
<th>Reflexive Perception</th>
<th>Meta-Management</th>
<th>Management Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliberative Perception</td>
<td>Planning, Reasoning</td>
<td>Deliberative Action</td>
</tr>
<tr>
<td>Reactive Perception</td>
<td>Reflexes</td>
<td>Reflexive Action</td>
</tr>
</tbody>
</table>
Conceptual Analysis: HCogAff (Sloman 2001)
Components for Cognitive AI

- Whole, testable architectures

PSI theory

*Principles of Synthetic Intelligence*
(Dörner 1999; Bach 2003, 2009)