Themes, Challenges, Programs:
Summary of Workshop on Human-like Learning

Paul Cohen
Richness and Compression in Mental Representations

• Human sensori-motor/cognitive experience is *enormously* rich, and AI representations are comparatively poor and empty, and this is one reason we don't yet have AI

• Semantics, generalization, compression, gist, symbolic representation, and parsimonious models, plus the fact that we handle relatively few chunks at a time, suggest rich sensory representations are somehow compressed or chunked

• Richness is necessary for learning, compression is necessary for perception/reasoning

• Developmentally: Rich (verbatim) => Compressed (gist)

• If we could make a machine do Rich => Compressed, it would help us achieve AI
Big, rich, complex world and rich experiences and learning opportunities

- NL, vision, social and physical interaction could provide access to the rich world and populate memory
- However, NL, vision, and social/physical interaction aren't possible without memory (e.g., no semantics in NL)
- Plus, this isn't a program in NL, vision, etc.
- So we have a bootstrapping problem
- But perhaps bootstrapping is the program
Challenge: Cognitive Bootstrap

• Grow semantic/episodic memory by embedding an agent in a rich world to which it has access
• Access cannot require NL, vision, physical/social interaction
• If these skills develop, fine, but the performance tasks have to do with recall, planning, problem solving…
• Gaming technology is outstripping AI
• Realistic, machine accessible environments are available now or soon
• Will they be rich and nuanced enough (Geary)? They are developing faster than AI, which is stalled
Theme: Gist vs. Verbatim

Because we're always looking for gists we don't always notice the details.
Theme: Gist vs. Verbatim

Because we're always looking for gists we don't always notice the details.
Theme: Gist

- What is gist and levels of gist?
- Arguably no machine has gist memory, but generalization and classification and schemas are common; are these gists or gist-making processes? why or why not?
- If AI had solved the problem of gist it would also have solved the problem of semantics; we would have representations that say *a lot* about the world, semantic locality in memory, etc.
Challenge: Superhuman Memory

- Human memory is an extreme point in design space, tradeoffs in the design space are revealed by framing effects, intrusions, false memories, etc.
- Superhuman memory capitalizes on computer architectures, networks, other nonhuman architectural features to provide gist-like memory without irrational (in SEU sense) effects
- Tests of superhuman memory:
  - Confidence doesn't change unless information is provided
  - Identify gist without losing verbatim detail
  - Represent in terms of gist but don't throw out specific details that affect probabilities (rational framing)
  - Represent in terms of gist without increased probability of false memory
  - Rational decision-making
Theme: Control, Mental Simulation, and General Cognition

- Prediction and control are ur-tasks
- They drive almost all machine learning
- The models they provide inform planning, problem-solving, robotics, information retrieval, and much else in AI
- Distinctions such as verbatim/gist are viewed as points on a spectrum when viewed as means or models for prediction and control
- Geary characterizes the drive to general intelligence as a drive to control
- The army characterizes success as control
- Control theory is almost irrelevant to psychological processes that produce prediction and control
- If one could pick only one task that would drive the development of general intelligence, and most of AI, it would be learning predictive models for the purpose of control
Predictive models for control

• Kalman filters
• RL policies
• POMDPs
• Rule books at nuclear power plants
• Expert systems for medical triage

• What's missing from these is explanations, reasons, semantics

NAH!
Challenge: The Explaining Mind

- Learn predictive models for control (why must they be gist-like?)
- Adapt models from other domains (what features of the models support this transfer?)
- Explain the models and one's actions based (the semantic part)

Example: Machine learns to be a Sims character in a Sims game. It must explain its actions. Its models needn't be right (e.g., "Girls love this..."), but the machine must be able to explain them.
Theme: Information, encoding, gist and general intelligence (again)

This is a minimum-description length view: The model plus exceptions are the minimum-length description of the scene.
Theme: Information, encoding, gist and general intelligence (again)

This is a minimum description length view: The model plus exceptions are the minimum-length description of the scene.

The same principle might explain the relationship between verbatim-like models and gist-like models.
Challenge: Model the development or learning of gist-like representations

Instead of designing representations (what every AI engineer does) consider learning them. Questions we'd like to answer:

- If there are semantic/ representational primitives, why *these* not *those*
- Why does the mind form classes so inevitably, why do classes have the structures they have?
- Why do some syntactic configurations never become representations, even when they are just as predictive as others?
- Is prediction a sufficient task to drive learning representations?
- Is there anything to the image-schematic ideas in the cognitive development literature, can these ideas be implemented?
- In what sort of test environment can we model the development or learning of representations (i.e., what are the performance tasks and metrics?)
Challenge: Semantic Autonomy
The machine knows what formal objects mean

Current practice is to tell the machine as little about the meanings of symbols as possible (contra gist). In general, if syntactic operations on formal objects make sense to us, we don't worry whether they make sense to the machine. Consequently we are semantic babysitters for our machines, always checking whether their conclusions make sense, as they cannot do it themselves.

Two approaches to semantic autonomy:

1. We tell the machine vast amounts of things about headaches, meninges, infections, anatomy, life, death, medical practice, etc. Cyc, Ontosaurus...
2. The machine learns meanings of formal objects by and for itself – "Show the program how to find and fix a bug and the program will work forever" Oliver Selfridge
What I learned at the workshop

• The reason we don't have human cognitive or learning capabilities is *not* a lack of algorithms
• It might be a lack of access to a rich environment (see Bootstrapping challenge)
• It might be a focus on the wrong kinds of problems (see Control and Prediction challenge)
• It almost certainly is a lack of good ideas about semantics, memory, gist-like representations
• It almost certainly is related to accepting semantic babysitting instead of requiring semantic autonomy, gist-like memory, and providing environments in which it can develop