
mit
media
lab

Measuring the Progress of AI Benchmark Problems

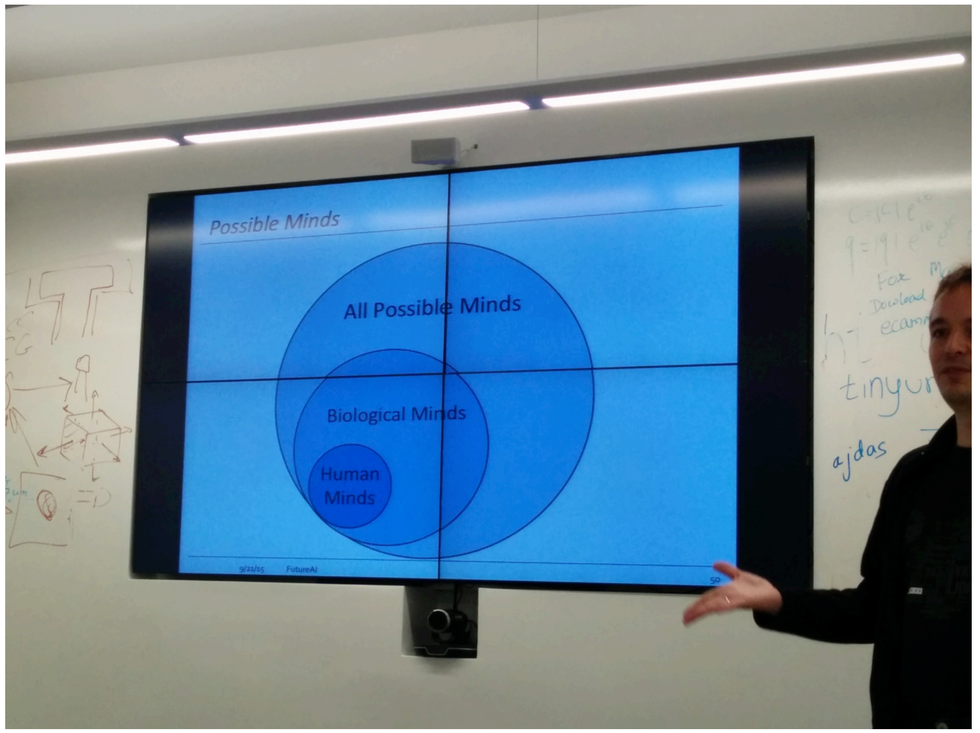
Dhaval Adjodah & Kane Hadley

Structure of presentation

- The Big Picture
- Fundamental problem: measuring AGI and it's progress
- Meta-problem and meta-solution
- Common Framework
- Bottom-up and top-down approaches
- Tasks
- ***DISCUSSION!!!!!!***

The big picture:

- The space of possible architectures of minds is huge
- We imagine minds through our own limited cognition
- There should be a more scientific/ general map of measuring our progress



The space of possible minds by Joscha Bach

Fundamental problem: measuring AGI

- We have a formal definition of general intelligence based on the Solomonoff-Levin distribution (Legg and Hutter 2007)
- Problems:
  - Not all intelligent systems are explicitly reward-seeking (Goertzel 2010)
  - The above definition is superficial in that it would pass the Norvig ‘test’ but not the Chomsky ‘test’
  - No way fundamentally to measure ‘being conscious’ since it is an ‘inside’ experience
  - There is really no way to prove anybody else is conscious

Alternate measurements:

- Collective intelligence (Woolley 2010)
  - Chimps beat human at individual intelligence tests, but never collective intelligence tests
- Distillation learning / Dark Knowledge (Hinton 2014)
- Emergent intelligence / Generalization measurement
- Disobedience
  - Internal goals

Meta-problem: collaborating on AGI research effectively

“But aside from the many technological and theoretical challenges involved in this effort, we feel the greatest impediment to progress is the absence of a common framework for collaboration and comparison of results”

“A common goal and a shared understanding of the landscape ahead of us will be crucial to that success, and it was the aim of our workshop to make substantial progress in that direction”

Meta-solution: building a common framework

“A common goal and a shared understanding of the landscape ahead of us will be crucial to that success, and it was the aim of our workshop to make substantial progress in that direction”

“research paradigms could be used to spawn a slightly different requirements list, but we must start somewhere if we are to make progress as a community”

Meta-solution: building a common framework

“A common goal and a shared understanding of the landscape ahead of us will be crucial to that success, and it was the aim of our workshop to make substantial progress in that direction”

“research paradigms could be used to spawn a slightly different requirements list, but we must start somewhere if we are to make progress as a community”

“To test the capability of any AGI system, the characteristics of the intelligent agent and its assigned tasks within the context of a given environment must be well specified. Failure to do this may result in a convincing demonstration, but make it exceedingly difficult for other researchers to duplicate experiments and compare and contrast alternative approaches and implementations”

Common Framework Part 1: Characteristics

C1. The environment is complex, with diverse, interacting and richly structured objects.

C2. The environment is dynamic and open.

C3. Task-relevant regularities exist at multiple time scales.

C4. Other agents impact performance.

C5. Tasks can be complex, diverse and novel.

C6. Interactions between agent, environment and tasks are complex and limited.

C7. Computational resources of the agent are limited.

C8. Agent existence is long-term and continual.

The characteristics shown provide the necessary (if not sufficient) degrees of dynamism and complexity that will weed out most “narrow AI” approaches at the outset, while challenging researchers to continually con- sider the larger goal of AGI during their work on subsystems and distinct capabilities

Common Framework Part 2: Characteristics

R0. New tasks do not require re-programming of the agent

R1. Realize a symbol system

Represent and effectively use:

- R2. Modality-specific knowledge
- R3. Large bodies of diverse knowledge
- R4. Knowledge with different levels of generality
- R5. Diverse levels of knowledge
- R6. Beliefs independent of current perception
- R7. Rich, hierarchical control knowledge
- R8. Meta-cognitive knowledge

R9. Support a spectrum of bounded and unbounded deliberation

R10. Support diverse, comprehensive learning

R11 Support incremental, online learning

There are nearly as many different AGI architectures as there are researchers in the field. If we are ever to be able to compare and contrast systems, let alone integrate them, a common set of architectural features must form the basis for that comparison

Challenges:

- One challenge is to find tasks and environments where all of these characteristics are active, and thus all of the requirements must be confronted.

- A second challenge is that the existence of an architecture that achieves a subset of these requirements, does not guarantee that such an architecture can be extended to achieve other requirements while maintaining satisfaction of

- A third challenge, that of defining the landscape, the focus of the rest of this presentation

  - “we initially had neither a well defined starting point nor a commonly agreed upon target result”

Final destination

- The final destination, full human-level artificial general intelligence, encompasses a system that could learn, replicate (and possibly exceed) human level performance in the full breadth of cognitive and intellectual abilities

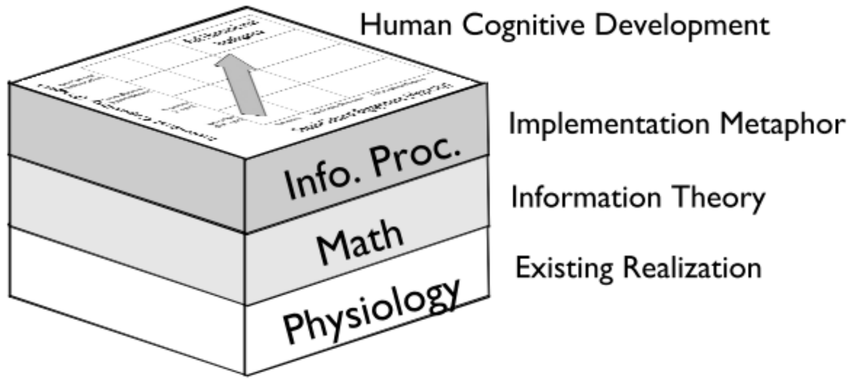
This is where the requirements from previous tables (characteristics and architectures of AGI) is important, in addition to ways to measure intelligence.

The harder problem: the starting point

- The starting point is more problematic, since there are many current approaches to achieving AGI that assume different initial states
- “We finally settled on a developmental approach to the roadmap, following human cognitive development from birth through adulthood”
- Two ways to think about it:
  - Top Down: looking at the emergent characteristics of AI (e.g. IQ)
  - Bottom-up: looking at how the mind is implemented (e.g. information theory)

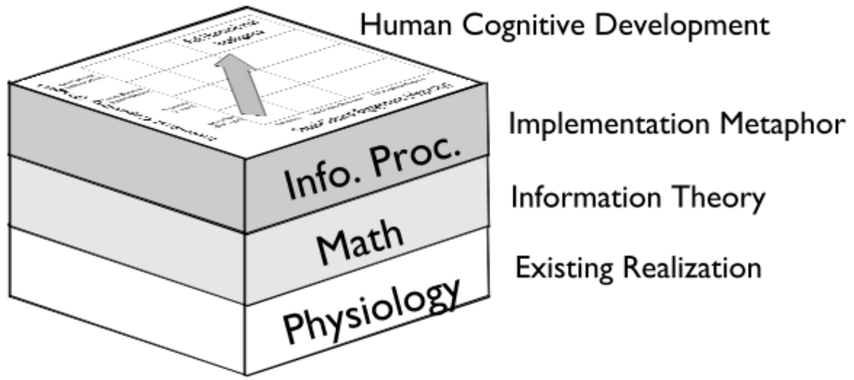
Bottom-up: The Psychological Perspective:

- Holistic perspective incorporating genetic, biochemical, and neural mechanisms, among others
- Consistent pattern across a wide range of cultures, physical environments, and historical time-periods
- Explains differences between typical and atypical patterns of development (e.g., autism, ADHD, learning disorders, disabilities, etc.)



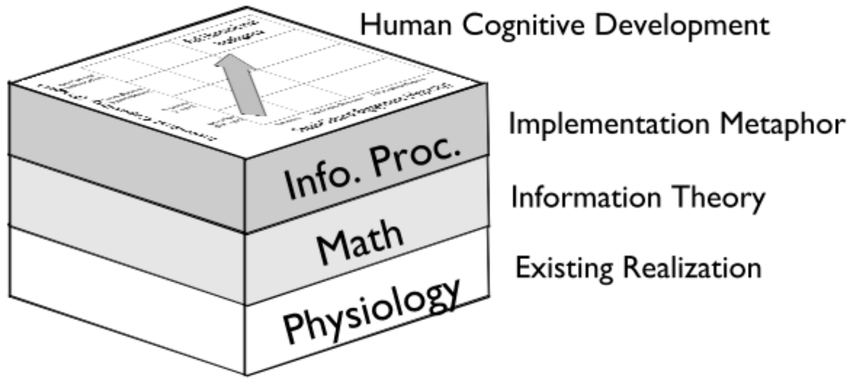
Bottom-up: The Mathematical Perspective

- Typified by formal definition of general intelligence based on the Solomonoff-Levin distribution (Legg and Hutter 2007)
- They define intelligence as the average reward-achieving capability of a system, calculated by averaging over all possible reward-summable environments



Bottom-up: The Information processing Perspective

- Provides a more direct mapping to the target implementation for AGI systems.
- Cognitive development in infancy and childhood is due to changes in both “hardware” and “software”
- Rather than advocating a stage-based approach, this perspective often highlights processes of change that are gradual or continuous.



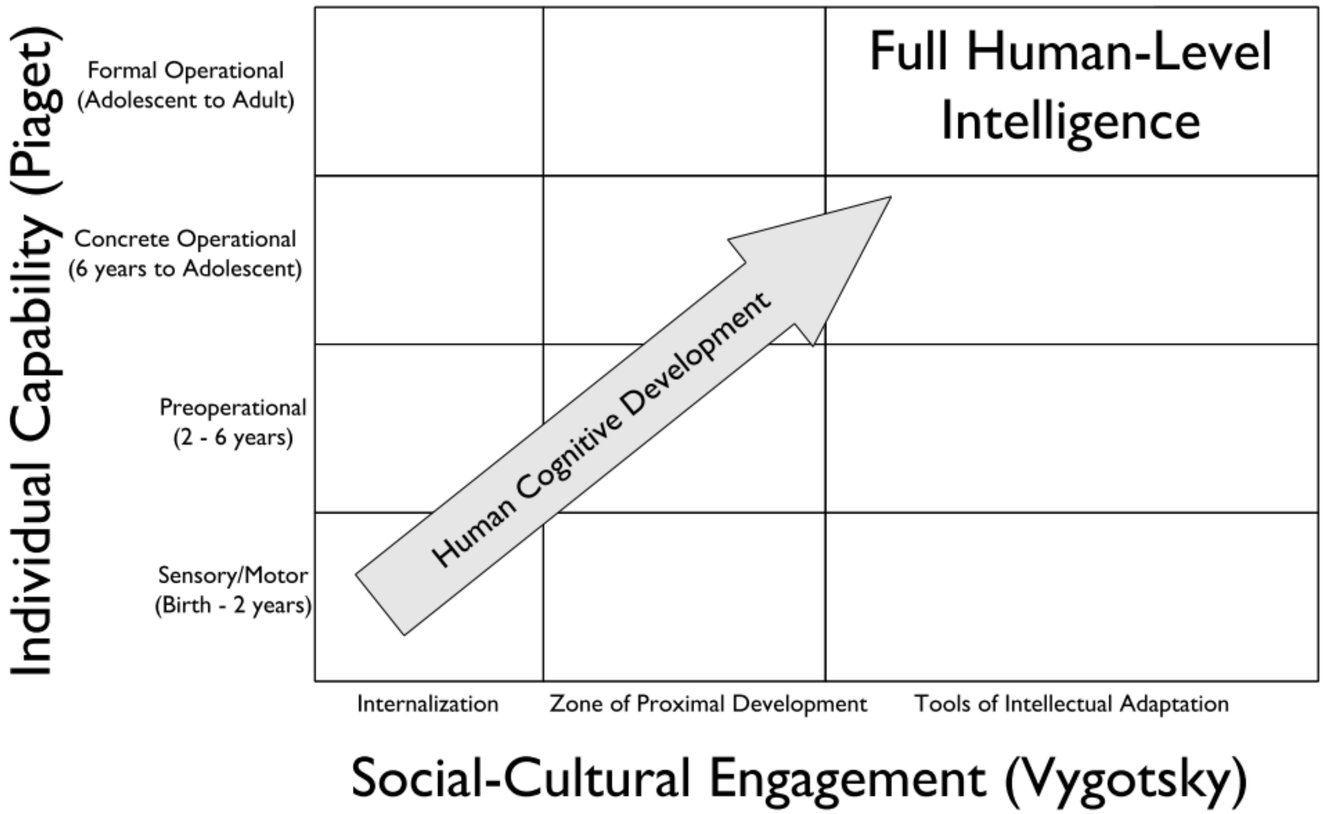
Top Down: Characterizing Human Cognitive Development

- The psychological approach to intelligence encompasses a broad variety of sub-approaches rather than presenting a unified perspective
  - Spearman psychological factor ***g*** is biologically determined, and represented the overall intellectual skill level of an individual
  - Binet and Simon scale that provided comprehensive age norms, so that each child could be systematically compared with others across age and intellectual level
  - Terman introduced the notion of an intelligence quotient or ***IQ***, which is computed by dividing the test-taker's mental age (i.e., their age-equivalent performance level) by their physical or chronological age
- Psychologists don't agree on intelligence as a single, undifferentiated capacity (fortunately!)

Top-down landscape:

- Researchers in the field of cognitive development seek to:
  - describe processes of intellectual change,
  - while identifying and explaining the underlying mechanisms (both biological and environmental) that make these changes possible.
- Contemporary theories of cognitive development are very diverse and defy simple systematization
- Two major schools of thought, those of Piaget and Vygotsky, will serve as axes for our AGI landscape.

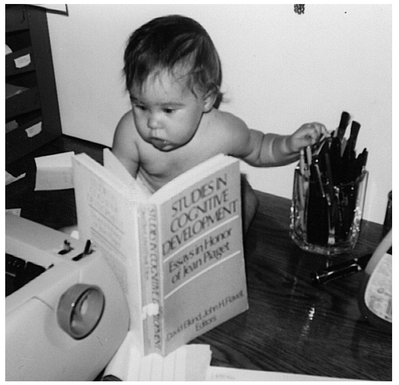
Top-Down Landscape



Landscape of Human Cognitive Development

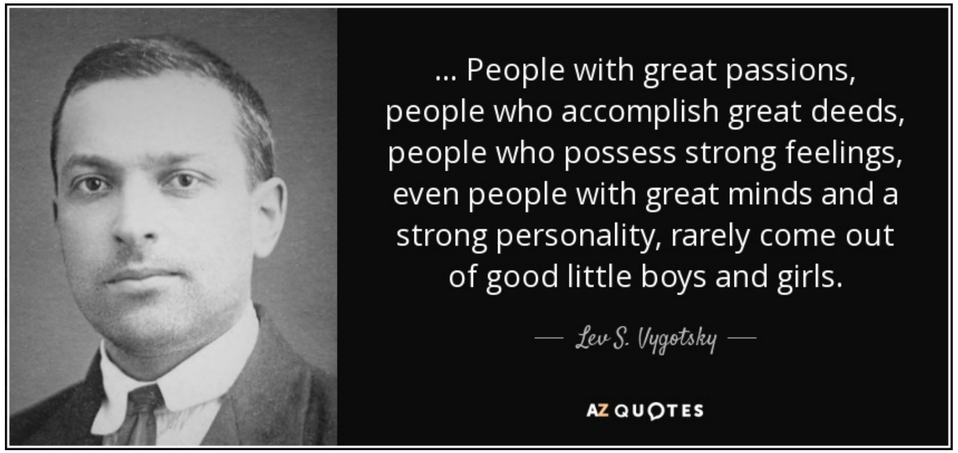
Piaget

- Piaget proposed that humans develop through four qualitatively-distinct stages of development
  - sensorimotor stage (0-2 years): infants acquire a rich repertoire of perceptual and motor skills
  - infants enter the preoperational stage (2-6 years) as they acquire the capacity to mentally represent their experiences
  - next stage (6 years to adolescence): children at the concrete operational level master basic elements of logical and mathematical thought
  - final stage of development, formal operations, begins in adolescence and includes the use of deductive logic, combinatorial reasoning, and the ability to reason about hypothetical events



Vygotsky

- His classic theory of cognitive development emphasizes the sociocultural perspective (Vygotsky 1986)
  - the capacity for thought begins by acquiring speech (i.e., thinking “out loud”) which gradually becomes covert or internalized
  - parents, teachers, and skilled peers facilitate development by helping the child function at a level just beyond what they are capable of doing alone
  - unique set of objects, ideas, and traditions that guide learning. These tools of intellectual adaptation not only influence the pattern of cognitive development, but also serve as constraints on the rate and extent of development



Identifying Tasks

Environment

Embodiment

C1. The environment is complex, with diverse, interacting and richly structured objects.

C2. The environment is dynamic and open.

C3. Task-relevant regularities exist at multiple time scales.

C4. Other agents impact performance.

C5. Tasks can be complex, diverse and novel.

C6. Interactions between agent, environment and tasks are complex and limited.

C7. Computational resources of the agent are limited.

C8. Agent existence is long-term and continual.

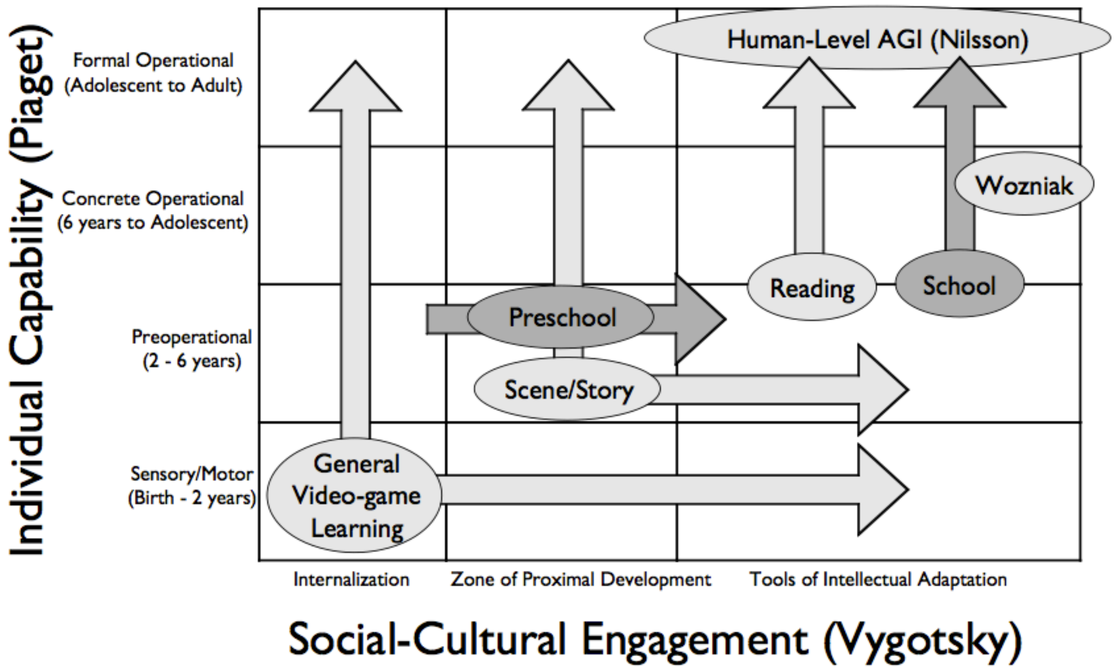
Identifying Tasks

- General Video-game Learning
- Preschool Learning
- Reading Comprehension
- Story / Scene Comprehension
- School Learning
- The Employment Test

Roadmap

Environment

Embodiment



Additional Challenges

- Aesthetic Appreciation and Performance
- Structured Social Interaction
- Skills the require high cognitive function and integration

Discussion

- All AI tests so far are testing the desired emergent behaviors of the architecture, but there are no tests for deeper cognition such as:

  •Developing internal goals different from the environmental norm
  •Introspection
  •Learning / teaching / generalization
  •Wisdom / knowledge

Discussion

- What other approach should we take than following human cognitive development from birth to adulthood?

Discussion

- Problem solving vs. Environment Utility
- Cognition as intelligence vs. Perception as intelligence

Discussion – Stanley Parable

Discussion – Monument Valley