

# Mathematical definition of “intelligence”

after

***Mathematical definition of “intelligence”  
(and consequences)***

by

Warren D. Smith, 2006

# intelligence = learning from experience

- many approaches model intelligent agents after sequential decision theory
  - Russell & Norwig: „intelligence is about agents in environments”
  - Hutter derives „an optimal rational reinforcement learning agent, embedded in an unknown environment”
  - Smith defines intelligence by a test, where the agent is to (learn to) solve NP problems knowing only scores for its answers
  - i.e. reinforcement can be as minimal as binary „good / bad”

# Why theory about general intelligence

Affordable-price computers will match processing power of human brain sometimes between now (Smith) and 2030 (Kurzweil), but computers are „not smart” yet.

- Smith's „dimensions of intelligence”:
  - generality: „willing to investigate any mental question”
  - asymptoticity: not initial competence, but asymptotic final competence
  - efficiency: „polynomial time and space algorithm”
- Pei Wang defines intelligence as
  - „the ability to adapt to the environment and to work with insufficient knowledge and resources”

# definition of intelligence [1]

- a usefully intelligent entity is one that can supply good “answers” to “questions” (both are general bitstrings),
- we recognize that only NP questions and P-verifiable answers are needed,
- we argue that any entity that cannot score well on certain such tests is not intelligent, while any entity that can score optimally well on such a test is extremely intelligent
- entity which does score optimally well on every such intelligence test exists

# definition of intelligence

- Beyond NP, we would need to test adversarial intelligences on e.g. General Game Playing (PSPACE).
- It is important to be able to solve problems that do not have (or do not have obvious) descriptions and definitions at all: in real life, often a large part of solving the problem first is to find a good problem statement.
- The score of an agent is measured as a function of time (not of the number of problem instances).

# What should Aiers do now?

„So we hereby propose that AI researchers instead concentrate on (1) achieving real understanding (i.e. **prove theorems**) and/or, (2) when doing experimental work with computers not amenable to complete theoretical understanding, to **make reproducible and standardized measurements** of Progress toward the Main Goal. We claim that both of these now are possible.”

# Uniformly Asymptotically Competitive Intelligence

- „We construct a UACI, asymptotically as intelligent as any other entity on every intelligence test simultaneously. Unfortunately this UACI consumes time exponential in the codelength of the competitor entity, but that is shown to be unavoidable.”
- The UACI enumerates all polytime algorithms, spends 50% time on evaluating their cumulative expected scores and 50% time on using the best one found so far.
- (But I don't know if the UACI also searches for the polytime tester – it seems so from description.)
- If better than  $\text{EXP}(\text{codelength}) = \text{“breaking the AES secret key cryptosystem”}$ .

# Previous Universality Results

- **Levin search**: inverts a given function  $g$  by spending  $2^{(-l)}$  time on every program of length  $l$  (thus with  $2^l$  slowdown to the optimal  $g^{(-1)}$  algo)
- Hutter improves that by also optimizing the algorithm for  $g$ .
- Schmidhuber's „**Goedel machine**“: runs a program to interact with an environment, simultaneously searching for another programs, and for proofs, and replaces current program with provably better one (giving higher expected future reward) when such found.

# Comparison with Marcus Hutter's

- Solomonoff's universal induction is more general: all algorithms (vs. P), but less general: exact answers vs. score feedback
  - NP questions and P answers both adequate and the only efficient IQ test
- AIXI is even more general: active environment (so a real agent framework)
  - „AI that tries to outthink you and predict your future actions, and intentionally tries to give you worse answers with smaller scores if it believes that that will influence you to keep awarding those scores for longer.”

# Comparison with Marcus Hutter's

- UACI „the lowest point” on Hutter's hierarchy of intelligences.
- Hutter's „universal intelligence test” vs. the claim that „there is no such thing as universal intelligence test”.
- Hutter's agents predict with an unbounded horizon (i.e. with harmonic reward discount: with effective horizon linear in the life of agent), Smith's have bounded horizon.

# Sources

[1] *Mathematical definition of “intelligence” (and consequences)*, Warren D. Smith, 2006