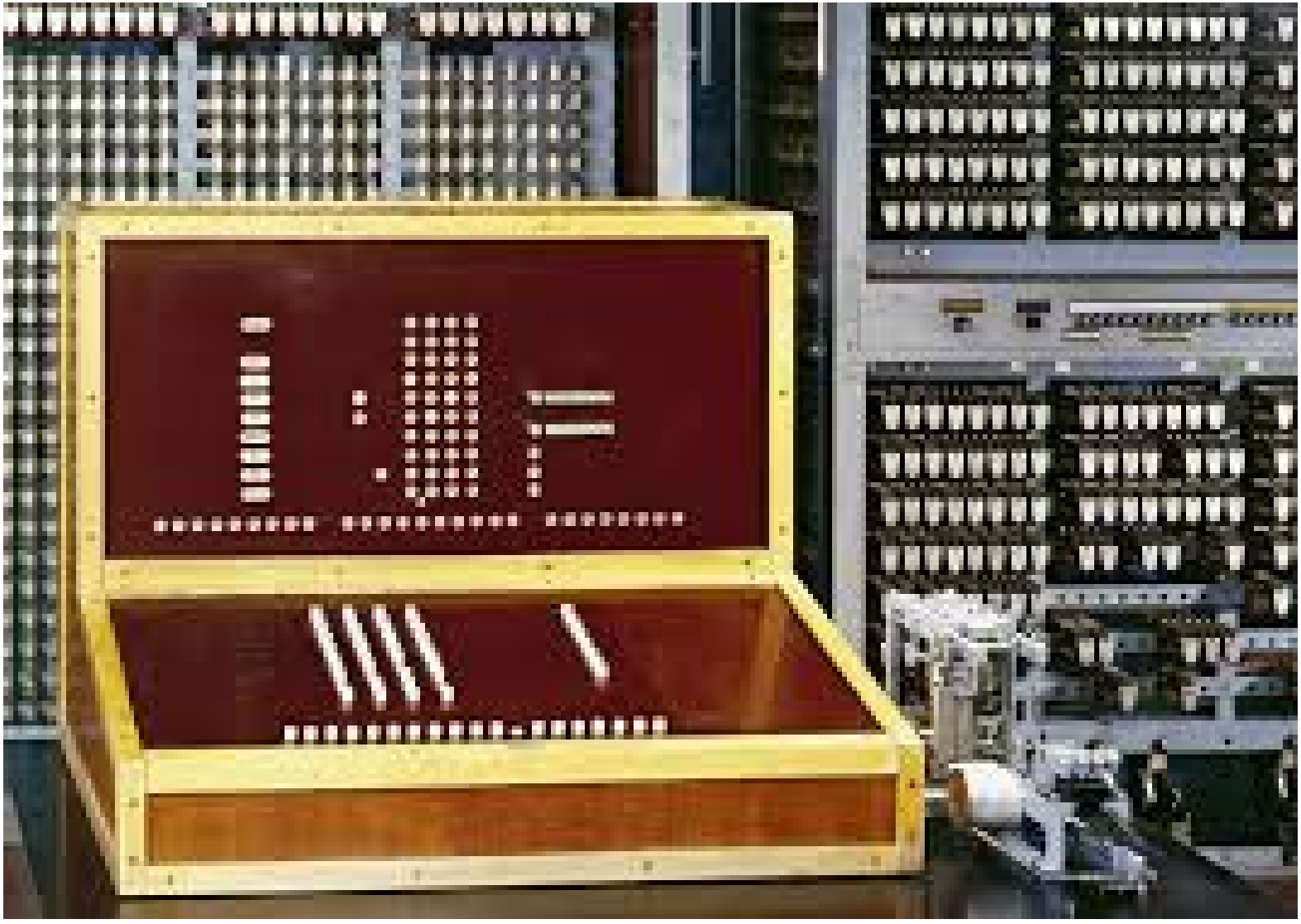
Intelligence as Inference or Forcing Occam on the World

Peter Sunehag and Marcus Hutter

2014 at AGI

Building a machine that simplifies the world



General Reinforcement Learning

- An environment $\nu(h_t, a_t) = (o_t, r_t)$ where $h_t = a_1 o_1 r_1, \ldots, a_t o_t r_t$.
- Maximize the discounted reward sum (return) $\sum_{i=t}^{\infty} r_i \gamma^{t-i}$ where $\gamma \in (0, 1)$
- A policy is a function $\pi(h_t) = a_t$
- $V_\nu^\pi(h_t) =$ expected return (in ν) following policy π after h_t
- The strictly rational policies are

$$\cup_\xi \arg\max_\pi V_\xi^\pi(h_t)$$

- The AIXI agent defines ξ as a mixture $\sum w_\nu \nu$ over all lower semi-computable environments and $\omega_\nu > 0$ for all such ν

Solving the problems with UAI

- Our previous work aimed at alleviating one main problem with Universal AI, the choice of reference machine, by picking finitely many (AGI'2012) or even an infinite exhausting sequence (AGI'2013).
- In these works where one picks the most optimistic ξ introduced so far, the agent is in a sense learning a machine (an optimistic one).
- Any agent is super-intelligent for some machine. Occam's razor CANNOT be true as a proposition about the world for all machines!. Imperative: Learn one for which Occam is true
- Second problem: Performing the optimization. Its incomputable.
- Idea: Solve both simultaneously by constructing a machine for which good programs are short.
- Good can both be in the sense of rewarding policies/agents and in the sense of implementing plausible environments
- The machine encodes our beliefs about the world and what is a high-achieving agent

What is a good machine

- Assume that there is a return function $R(p) \in \mathbb{R}_+$
- The return function is not explicitly available but is evaluated by running the program.
- R can depend on running speed. Hence the task is practical, though R could also be an AIXI objective in theory.
- Finding a p that results in a large $R(p)$ is the goal and for some R this is developing AGI (or UAI for an AIXI objective).
- We search on programs by coin-flipping bits until we got a program that runs and then we observe $R(p)$.
- We want a machine with high expected return $E(U) = \sum 2^{-\ell_U(p)} R(p)$.
- If $\sum 2^{-\ell_U(p)} \ell_U(p) R(p) > \sum 2^{-\ell_U(p)} \ell'_U(p) R(p)$, then (we prove) $E(U') > E(U)$. Approximate with set $\{p_i\}_{i=1}^{M}$ of programs on U.
- Stronger correlation between short length and high return gives a better machine. Evaluation without sampling/running programs on a new proposed machine. Only translation is required.

Building a world with Occam

- Development of more suitable reference machines is already happening by people who might not care about AGI or AI.
- New programming languages (and improvements with libraries) are developed where rewarding programs can be written shorter.
- Science and languages create concepts that simplify the world.

- Our world is constructed by Occam's imperative, simpler is better.
- Large number of tasks has been simplified for a long time. Restaurants simplify cooking to order and pay.

Do we need to intentionally build AGI

- Will AGI just emerge from the human activity of simplifying the world?
- Will the internet just increasingly "wake up" as suggested by Goertzel (and others)?
- Perhaps, but

- A) we need to be able to perform the search/development of the program(s) on top though this might happen in a hierarchy of meta-services
- B) It would be a very static AGI that does not take over the improvement of the reference framework
- We make these two tasks into one here.

Conclusions

- Imperative: Develop reference machine that simplify the world
- Inspired by the "Planning as inference" paradigm we suggest to look for a machine on which good agents and good environment models have short implementation. The machine encodes what has been learnt so far.
- The process is already ongoing world-wide throughout history. The pace increases as the world gets simpler and more successful behaviors to be simplified are found since useful exploration gets easier (see also accommodation in evolutionary theory).
- Current work, using continuously parameterized functions like auto-encoders as the reference "machine" in a reward-modulated inference framework.
- Connection to neuroscience through reward-modulated spike-timing plasticity models and to psychology through the law of effect.