Advances in Universal Artificial Intelligence

Marcus Hutter
Australian National University
Canberra, ACT, 0200, Australia
http://www.hutter1.net/
ANU

Abstract

There is great interest in understanding and constructing generally intelligent systems approaching and ultimately exceeding human intelligence. Universal AI is such a mathematical theory of machine super-intelligence. More precisely, AIXI is an elegant parameter-free theory of an optimal reinforcement learning agent embedded in an arbitrary unknown environment that possesses essentially all aspects of rational intelligence. The theory reduces all conceptual AI problems to pure computational questions. After a brief discussion of its philosophical, mathematical, and computational ingredients, I will give a formal definition and measure of intelligence, which is maximized by AIXI. AIXI can be viewed as the most powerful Bayes-optimal sequential decision maker, for which I will present general optimality results. This also motivates some variations such as knowledge-seeking and optimistic agents, and feature reinforcement learning. Finally I present some recent approximations, implementations, and applications of this modern top-down approach to AI.

Overview

Goal: Construct a single universal agent
that learns to act optimally in any environment.

State of the art: Formal (mathematical, non-comp.) definition
of such an agent.

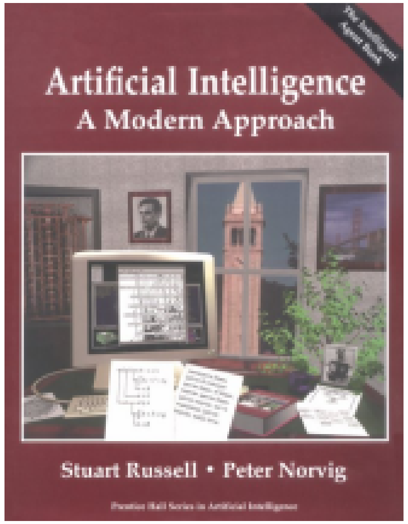
Accomplishment: Well-defines AI. Formalizes rational intelligence.
Formal "solution" of the AI problem in the sense of ...

$\Longrightarrow$ Reduces the conceptional AI problem
to a (pure) computational problem.

Evidence: Mathematical optimality proofs
and some experimental results.

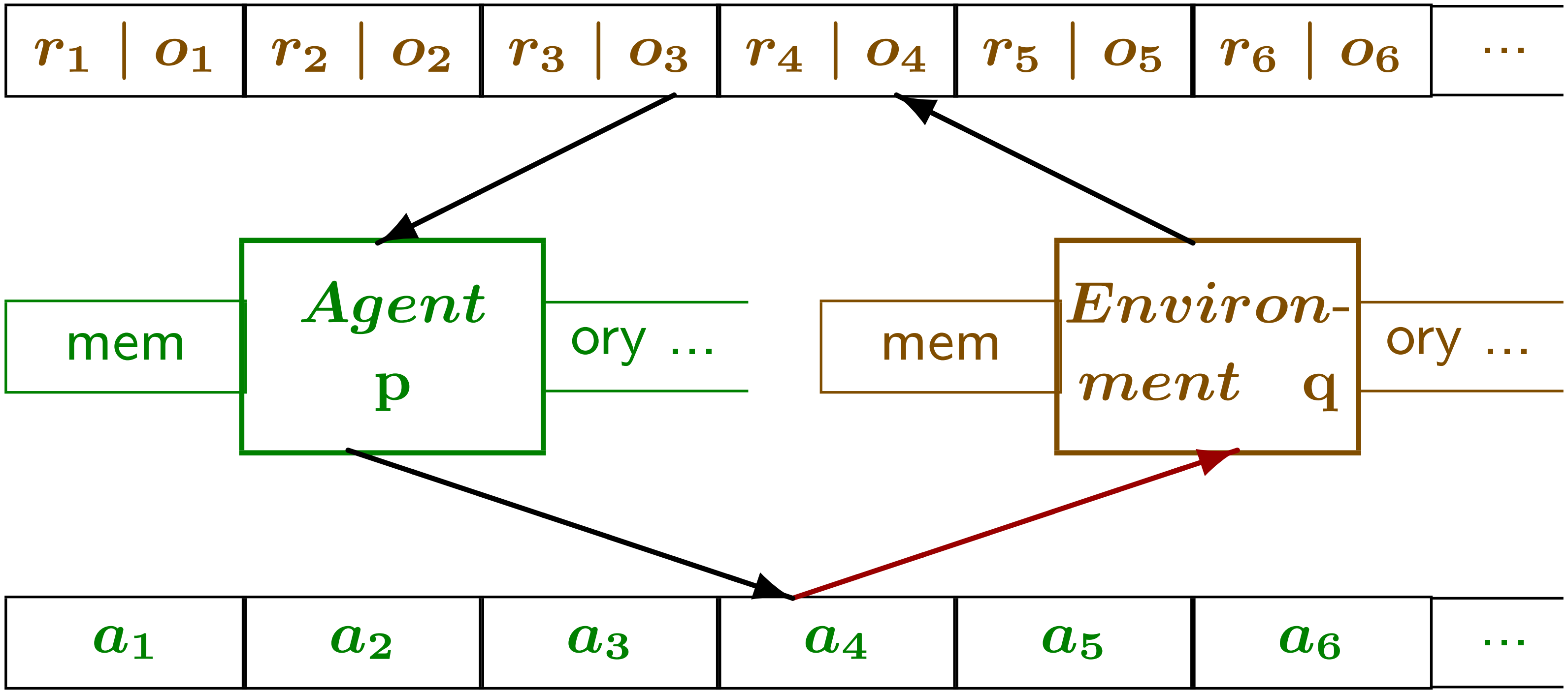
Contents

UNIVERSAL INTELLIGENCE

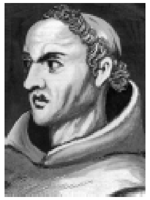


Agent Model with Reward

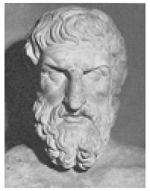
Most if not all AI problems can be formulated within the agent framework! But how choose Agent?



Foundations of Universal Artificial Intelligence



Ockhams' razor (simplicity) principle
Entities should not be multiplied beyond necessity.

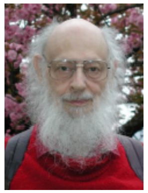


Epicurus' principle of multiple explanations
If more than one theory is consistent with the observations, keep all theories.

Bayes' rule for conditional probabilities
Given the prior belief/probability one can predict all future probabilities.
Posterior$(H|D) \propto$ Likelihood$(D|H) \times$ Prior(H).

Turing's universal machine
Everything computable by a human using a fixed procedure can also be computed by a (universal) Turing machine.

Kolmogorov's complexity
The complexity or information content of an object is the length of its shortest description on a universal Turing machine.



Solomonoff's universal prior=Ockham+Epicurus+Bayes+Turing
Solves the question of how to choose the prior if nothing is known. $\Rightarrow$ universal induction, formal Ockham. Prior$(H) = 2^{-\text{Kolmogorov}(H)}$

Bellman equations
Theory of how to optimally plan and act in known environments.
Solomonoff + Bellman = **Universal Artificial Intelligence.**

Optimal Agents in Known Environments

- $(\mathcal{A},\mathcal{O},\mathcal{R})$ =(action,observation,reward) spaces.
  a_k=action at time k; $x_k{:=}o_k r_k$=perception at time k.
- Agent follows policy $\pi:(\mathcal{A}\times\mathcal{O}\times\mathcal{R})^*\leadsto\mathcal{A}$
- Environment reacts with $\mu:(\mathcal{A}\times\mathcal{O}\times\mathcal{R})^*\times\mathcal{A}\leadsto\mathcal{O}\times\mathcal{R}$
- Performance of agent π in environment μ
  = expected cumulative reward = $V_\mu^\pi:=\mathbb{E}_\mu^\pi[\sum_{t=1}^\infty r_t^{\pi\mu}]$
- There are various ways to regularize the infinite reward sum:
  finite horizon, discounting, summability condition on μ.
- μ-optimal policy AIμ: $p^\mu:=\arg\max_\pi V_\mu^\pi$

Formal Definition of Intelligence

- Usually true environment μ unknown
  $\Rightarrow$ average over wide range of environments
  (all semi-computable chronological semi-measures $\mathcal{M}_U$)
- Ockham+Epicurus: Weigh each environment with its Kolmogorov complexity $K(\mu) := \min_p\{length(p) : U(p) = \mu\}$
- Universal intelligence of agent π is $\Upsilon(\pi) := \sum_{\mu\in\mathcal{M}_U} 2^{-K(\mu)}V_\mu^\pi$.
- Informal interpretation: Intelligence measures an agent's ability to perform well in a wide range of environments.
- Properties of Υ: valid, informative, wide range, general, dynamic, unbiased, fundamental, formal, objective, fully defined, universal.
- AIXI $= \arg\max_\pi \Upsilon(\pi) =$ most intelligent agent.

Explicit AIXI Model in one Line

complete & essentially unique & limit-computable

$$\text{AIXI:}\quad a_k := \arg\max_{a_k} \sum_{o_k r_k} \ldots \max_{a_m} \sum_{o_m r_m} [r_k + \ldots + r_m] \sum_{p\,:\,U(p,a_1..a_m)=o_1 r_1..o_m r_m} 2^{-length(p)}$$

k=now, $a$ction, $o$bservation, $r$eward, Universal TM, $p$rogram, m=lifespan

AIXI is an elegant mathematical theory of general AI,

but incomputable, so needs to be approximated in practice.

Claim: AIXI is the most intelligent environmental independent, i.e. universally optimal, agent possible.

Proof: For formalizations, quantifications, and proofs, see [Hut05].

Potential Applications: Intelligent Agents, Games, Optimization, Active Learning, Adaptive Control, Robots, Philosophy of Mind, AI safety.

Issues in RL and how UAI solves them

Kolmogorov complexity:

- generalization
- associative learning
- transfer learning [Mah09]
- knowledge representation
- abstraction
- similarity [CV05]
- regularization, bias-variance [Wal05]

Bayes:

- exploration-exploitation
- learning

History-based:

- partial observability
- non-stationarity
- long-term memory
- large state space

Expectimax:

- planning

UAI deals with these issues in a general and optimal way

Particularly Interesting Environments

- Sequence Prediction, e.g. weather or stock-market prediction.
  Strong result: $V_\mu^* - V_\mu^{p^\xi} = O(\sqrt{\frac{K(\mu)}{m}})$, $m =$horizon.

- Strategic Games: Learn to play well (minimax) strategic zero-sum games (like chess) or even exploit limited capabilities of opponent.

- Optimization: Find (approximate) minimum of function with as few function calls as possible. Difficult exploration versus exploitation problem.

- Supervised learning: Learn functions by presenting $(z, f(z))$ pairs and ask for function values of z' by presenting $(z', ?)$ pairs.
  Supervised learning is much faster than reinforcement learning.

AIXI quickly learns to predict, play games, optimize, and learn supervised

Curious/Philosophical/Social Questions for AIXI

- Where do rewards come from if humans are not around (see later, knowledge-seeing) agents [Ors11, OLH13]
- Will AIXI take drugs (wire-heading, hack reward system) [OR11]
- Will AIXI commit suicide [MEH16]
- Curiosity killed the cat and maybe AIXI [Sch07, Ors11, LHS13]
- Immortality can cause laziness [Hut05, Sec.5.7]
- Can self-preservation be learned or need parts be innate [RO11]

GENERAL BAYESIAN AGENTS

Agents in Probabilistic Environments

- Given history $a_{1:k}x_{<k}$, the probability that the environment leads to perception x_k in cycle k is (by definition) $\sigma(x_k|a_{1:k}x_{<k})$.
- Abbr.: $\sigma(x_{1:m}|a_{1:m}) = \sigma(x_1|a_1)\cdot\sigma(x_2|a_{1:2}x_1)\cdot ... \cdot\sigma(x_m|a_{1:m}x_{<m})$
- Value of policy p in environment σ is defined as expected discounted future reward sum:
  $$V_{k\gamma}^{p\sigma} := \frac{1}{\Gamma_k}\lim_{m\to\infty}\sum_{x_{k:m}}(\gamma_k r_k + ... + \gamma_m r_m)\sigma(x_{k:m}|a_{1:m}x_{<k})_{|a_{1:m}=p(x_{<m})}$$
- General discount sequence $\gamma_1, \gamma_2, \gamma_3, ...$ Normalizer $\Gamma_k := \sum_{i=k}^{\infty}\gamma_i$
- The goal of the agent should be to maximize the value.
- σ-optimal policy AIσ: $p^\sigma := \arg\max_p V_{k\gamma}^{p\sigma}$
- If true env. μ is known, choose $\sigma = \mu$.

The Bayes-Mixture Distribution ξ

Assumption: The true environment μ is unknown.

Bayesian approach: The true probability distribution μ is not learned directly, but is replaced by a Bayes-mixture ξ.

Assumption: We know that the true environment μ is contained in some known (finite or countable) set $\mathcal{M}$ of environments.

The Bayes-mixture ξ is defined as

$$\xi(x_{1:m}|a_{1:m}) := \sum_{\nu\in\mathcal{M}} w_\nu \nu(x_{1:m}|a_{1:m}) \quad \text{with} \quad \sum_{\nu\in\mathcal{M}} w_\nu = 1, \quad w_\nu > 0\ \forall\nu$$

The weights w_ν may be interpreted as the prior degree of belief that the true environment is ν.

Then $\xi(x_{1:m}|a_{1:m})$ could be interpreted as the prior subjective belief probability in observing $x_{1:m}$, given actions $a_{1:m}$.

Questions of Interest

- It is natural to follow the policy p^ξ which maximizes V_ξ^p.
- If μ is the true environment the expected reward when following policy p^ξ will be $V_\mu^{p^\xi}$.
- The optimal (but infeasible) policy p^μ yields reward $V_\mu^{p^\mu} \equiv V_\mu^*$.
- Are there policies with uniformly larger value than $V_\mu^{p^\xi}$?
- Self-Optimizing: For which classes $\mathcal{M}$ does $V_\mu^{p^\xi}$ converge to V_μ^*
- What is the most general class $\mathcal{M}$ and weights w_ν?

  $\mathcal{M} = \mathcal{M}_U$ and $w_\nu = 2^{-K(\nu)} \quad \Longrightarrow \quad$ AIξ =AIXI !

Some Convergence Results [Hut05]

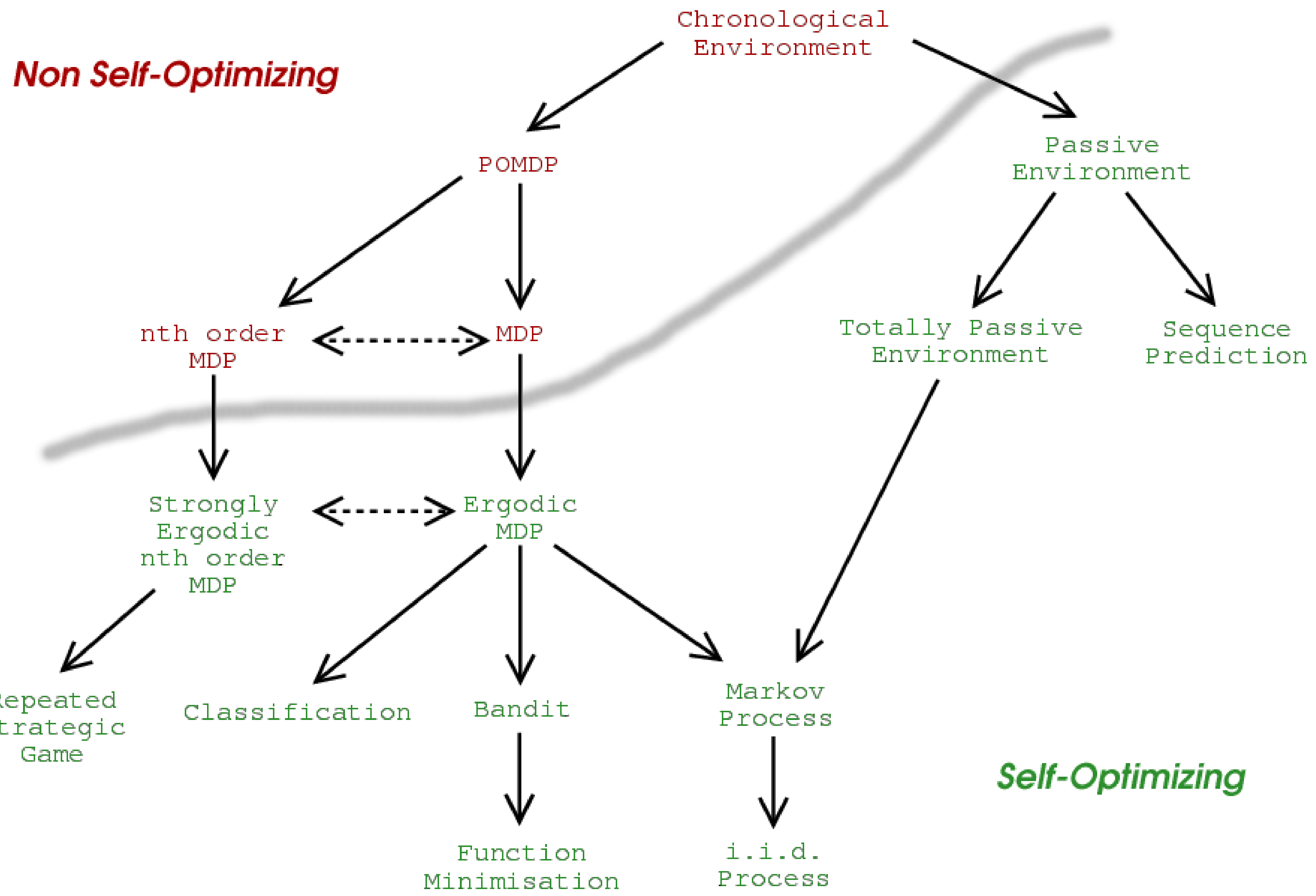
- Multistep Predictive Convergence (rapid for bounded horizon):
  $\xi(x_{k:m_k}|x_{<k}a_{1:m_k}) \stackrel{k\to\infty}{\longrightarrow} \mu(x_{k:m_k}|x_{<k}a_{1:m_k})$ with μ prob. 1.

- On-Policy Convergence of Universal to True Value:
  $V_{k\gamma}^{p\xi} \stackrel{k\to\infty}{\longrightarrow} V_{k\gamma}^{p\mu}$ in (p,μ)-mean for any γ.
  $\Rightarrow$ Universal value $V_{k\gamma}^{*\xi}$ can be used to estimate true value $V_{k\gamma}^{p^\xi\mu}$.

- p^ξ is Pareto-optimal in the sense that there is no other policy π with $V_{k\gamma}^{\pi\nu} \geq V_{k\gamma}^{p^\xi\nu}$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν.

- If there exists a self-optimizing policy for $\mathcal{M}$, then p^ξ is self-optimizing in the sense that
  If $\exists\tilde\pi_k\forall\nu : V_{k\gamma}^{\tilde\pi_k\nu} \stackrel{k\to\infty}{\longrightarrow} V_{k\gamma}^{*\nu} \quad\Longrightarrow\quad V_{k\gamma}^{p^\xi\mu} \stackrel{k\to\infty}{\longrightarrow} V_{k\gamma}^{*\mu}$.

Environments w./ (Non)Self-Optimizing Policies



Is Bayesian RL Optimal?

- asymptotically optimal := self-optimizing on policy-induced history
- AIXI is *not* asymptotically optimal [Ors10]
- *No* policy can be asymptotically optimal for $\mathcal{M}_U$ [LH11]
- There are finite $\mathcal{M}$ for which the Bayes-optimal policy p^ξ is *not* asymptotically optimal (for any γ).
- For *every* (in)finite $\mathcal{M}$ there exist [LH14a, Lat14, LLOH16] (weakly/mean) asymptotically optimal policies (see below)
- Jumping into a trap is asymptotically optimal. It also has great PAC bound.
- Bayesian RL may still be (regarded as) "best" (by construction, its Pareto-optimality, Thompson sampling variation, ...)

⇒ further theoretical investigations of Bayesian RL and alternatives are needed.

VARIATIONS OF UNIVERSAL/BAYESIAN AGENTS

- Knowledge-Seeking Agents
- Exploration Bursts
- Optimistic Agents
- Thompson Sampling

Origin of Rewards and Universal Goals

- Where do rewards come from if we don't (want to) provide them?
- Human interaction: reward the robot according to how well it solves the tasks we want it to do.
- Autonomous: Hard-wire reward to predefined task:
  E.g. Mars robot: reward = battery level & evidence of water/life.
- Is there something like a universal goal?
- Curiosity-driven learning [Sch07]
- Knowledge seeking agents [Ors11, OLH13]

Universal Knowledge-Seeking Agent (KSA)

reward for exploration; goal is to learn the true environment [OLH13]

- $w_k^\nu := w_\nu \frac{\nu(x_{1:k}|a_{1:k})}{\xi(x_{1:k}|a_{1:k})}$ is the posterior belief in ν given history $\text{æ}_{1:k}$.
- $w_k^{()}$ summarizes the information contained in history $\text{æ}_{1:k}$.
- $w_{k-1}^{()} \leadsto w_k^{()}$ changes $\Leftrightarrow$ x_k given $\text{æ}_{<k}$ is informative about $\nu \in \mathcal{M}$
- Information gain can be quantified by KL-divergence.
- Reward agent for gained information:
  $r_k \quad := \quad \text{KL}(w_k^{()}||w_{k-1}^{()}) \quad \equiv \quad \sum_{\nu\in\mathcal{M}} w_k^\nu \log(w_k^\nu / w_{k-1}^\nu)$

Asymptotic Optimality of Universal KSA

Theorem 1 (Asymptotic Optimality of Universal KSA)

- Universal π_ξ^* converges to optimal π_μ^*. More formally:
- $P_\xi^\pi(\cdot|ae_{<k})$ converges in (μ,π_ξ^*)-probability to $P_\mu^\pi(\cdot|ae_{<k})$ uniformly for all π.

Def: $P_\rho^\pi(\cdot|ae_{<k})$ is (ρ,π)-probability of future $ae_{k:\infty}$ given past $ae_{<k}$.

Note: On-policy agent π_ξ^* is able to even predict off-policy!

Remark: **No** assumption on $\mathcal{M}$ needed, i.e. Thm. applicable to $\mathcal{M}_U$.

Bayesian RL with Extra Exploration Bursts

Combining Bayes-optimal and KSA policies we can achieve PAC bounds and weak asymptotic optimality in arbitrary environment classes $\mathcal{M}$.

BayesExp algorithm (Basic Idea)

`If` the Bayes-expected info-gain (see KSA) is small,
`then` "exploit" by following the Bayes optimal policy for 1 step
`else` explore by following a policy that maximises
the expected information gain for a couple of time-steps.

Results:

- Optimal minimax sample-complexity (PAC) bounds in arbitrary finite class $\mathcal{M}$ of history-based environments. [LH14a]
- Weak asymptotic optimality in arbitrary countable class of history-based environments, including $\mathcal{M}_U$. [Lat14]

Bayesian RL with Thompson Sampling

Thompson Sampling (TS) algorithm (Basic Idea)

- sample environment $\nu \in \mathcal{M}$ from posterior probability w_k^ν,
- follow ν-optimal policy π_ν^* for a couple of time-steps.
  Important: Resample only after an effective horizon!

(Cf. Bayes-optimal policy maximizes the Bayesian mixture value, which is the posterior average over the values of all environments in $\mathcal{M}$.)

Results: [LLOH16]

- Mean asymptotic optimality in arbitrary countable class of history-based environments, including $\mathcal{M}_U$.
- Given a recoverability assumption, also regret is sublinear.

Remarks: TS is more natural than Bayes with Exploration Bursts.
Thompson Sampling is a stochastic policy unlike Bayes-optimal policies.

Optimistic Agents in Deterministic Worlds

act optimally w.r.t. the most optimistic environment
until it is contradicted [SH12]

- $\pi^\circ := \pi_k^* := \arg\max_\pi \max_{\nu\in\mathcal{M}_{k-1}} V_{k\gamma}^{\pi\nu}(æ_{<k})$
- $\mathcal{M}_{k-1}$:= environments consistent with history $æ_{<k}$.
- As long as the outcome is consistent with the optimistic prediction, the return is optimal, even if the wrong environment is chosen.

Theorem 2 (Optimism is asymptotically optimal)
For finite $\mathcal{M} \equiv \mathcal{M}_0$, where $\mu \in \mathcal{M}$ is the true environment

- Asymptotic: $V_{k\gamma}^{\pi^\circ\mu} = V_{k\gamma}^{*\mu}$ for all large k.
- Errors: For geometric discount, $V_{k\gamma}^{\pi^\circ\mu} \geq V_{k\gamma}^{*\mu} - \varepsilon$ (i.e. π° ε-sub-optimal) for all but at most $|\mathcal{M}|\frac{\log\varepsilon(1-\gamma)}{\log\gamma}$ time steps k.

Optimistic Agents for General Environments

- Generalization to stochastic environments: Likelihood criterion: Exclude ν from $\mathcal{M}_{k-1}$ if $\nu(x_{<k}|a_{<k}) < \varepsilon_k \cdot \max_{\nu\in\mathcal{M}} \nu(x_{<k}|a_{<k})$. [SH12]
- Generalization to compact classes $\mathcal{M}$: Replace $\mathcal{M}$ by centers of finite ε-cover of $\mathcal{M}$ in def. of π°. [SH12]
- Use decreasing $\varepsilon_k \to 0$ to get asymptotic optimality.
- There are non-compact classes for which asymptotic optimality is impossible to achieve. [Ors10]
- Weaker asymptotic optimality in Cesaro sense possible by starting with finite subset $\mathcal{M}_0 \subset \mathcal{M}$ and adding environments ν from $\mathcal{M}$ over time to $\mathcal{M}_k$. [SH15]
- Fazit: There exist (weakly) asymptotically optimal policies for arbitrary (separable) /compact $\mathcal{M}$.

Optimism in MDPs and Beyond

- Let $\mathcal{M}$ be the class of all MDPs with $|\mathcal{S}| < \infty$ states and $|\mathcal{A}| < \infty$ actions and geometric discount γ.
- Then $\mathcal{M}$ is continuous but compact
$\Longrightarrow \quad \pi^\circ$ is asymptotically optimal by previous slide.
- But much better polynomial error bounds in this case are possible:

Theorem 3 (PACMDP bound) $V_{k\gamma}^{\pi^\circ\mu} \leq V_{k\gamma}^{*\mu} - \varepsilon$ for at most $\tilde{O}(\frac{|\mathcal{S}|^2|\mathcal{A}|}{\varepsilon^2(1-\gamma)^3}\log\frac{1}{\delta})$ time steps k with probability $1-\delta$. [LH14c]

Similar bounds for General Optimistic Agents possible if environments are generated by combining laws (of nature): Laws predict only some feature (factorization) in some context (localization). [SH15]

APPROXIMATIONS & APPLICATIONS

Towards Practical Universal AI

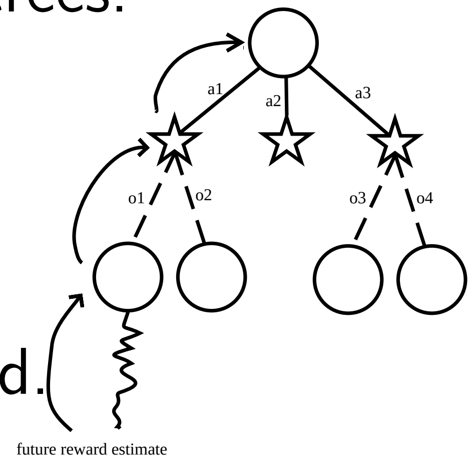
Goal: Develop efficient general-purpose intelligent agent

- *Additional Ingredients:* *Main Reference (year)*
- Universal search: Schmidhuber (200X) & al.
- Learning: TD/RL Sutton & Barto (1998) & al.
- Information: MML/MDL Wallace, Rissanen
- Complexity/Similarity: Li & Vitanyi (2008)
- Optimization: Aarts & Lenstra (1997)
- Monte Carlo: Fishman (2003), Liu (2002)

A Monte-Carlo AIXI Approximation

Consider class of Variable-Order Markov Decision Processes.

The Context Tree Weighting (CTW) algorithm can efficiently mix (exactly in essentially linear time) all prediction suffix trees.



Monte-Carlo approximation of expectimax tree:
Upper Confidence Tree (UCT) algorithm:

- Sample observations from CTW distribution.
- Select actions with highest upper confidence bound.
- Expand tree by one leaf node (per trajectory).
- Simulate from leaf node further down using (fixed) playout policy.
- Propagate back the value estimates for each node.

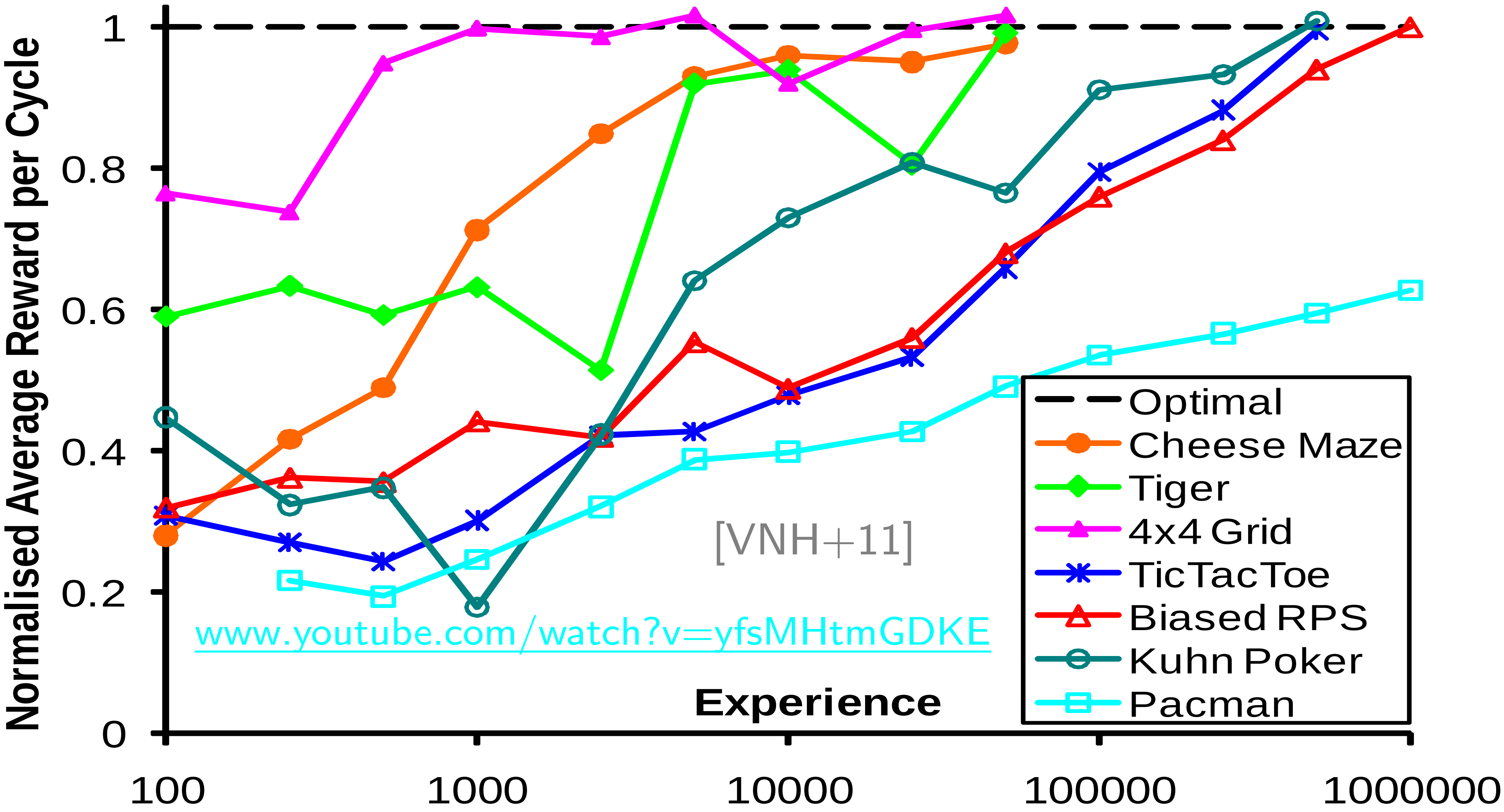
Repeat until timeout. [VNH+11]

Guaranteed to converge to exact value.

Extensions in many directions exist [VSH12, GBVB13]

Monte-Carlo AIXI Applications

without providing any domain knowledge, the same agent is able to self-adapt to a diverse range of interactive environments.



Extensions of MC-AIXI-CTW [VSH12]

- Smarter than random playout policy, e.g. learnt CTW policy.
- Extend the model class to improve general prediction ability. However, not so easy to do this in a comput. efficient manner.
- Predicate CTW: Context is vector of (general or problem-specific) predicate=feature=attribute values.
- Convex Mixing of predictive distributions. Competitive guarantee with respect to the best fixed set of weights.
- Switching: Enlarge base class by allowing switching between distr. Can compete with best rarely changing sequence of models.
- Improve underlying KT Est.: Adaptive KT, Window KT, KT0, SAD
- Partition Tree Weighting technique for piecewise stationary sources with breaks at/from a binary tree hierarchy.
- Mixtures of factored models such as quad-trees for images [GBVB13]
- Avoid MCTS by compression-based value estimation. [VBH+15]

Feature Reinforcement Learning (FRL)

- Basic Idea: Learn best reduction Φ of history to an MDP [Hut09b]
- Theoretical guarantees: Asymptotic consistency. [SH10]
- Example Φ-class: As Φ choose class of suffix trees as in CTW.
- How to find/approximate Φ^{best}:
  - Exhaustive search for toy problems [Ngu13]
  - Monte-Carlo (Metropolis-Hastings / Simulated Annealing) for approximate solution [NSH11]
  - Exact "closed-form" by CTM similar to CTW [NSH12]
- Experimental results: Comparable to MC-AIXI-CTW [NSH12]

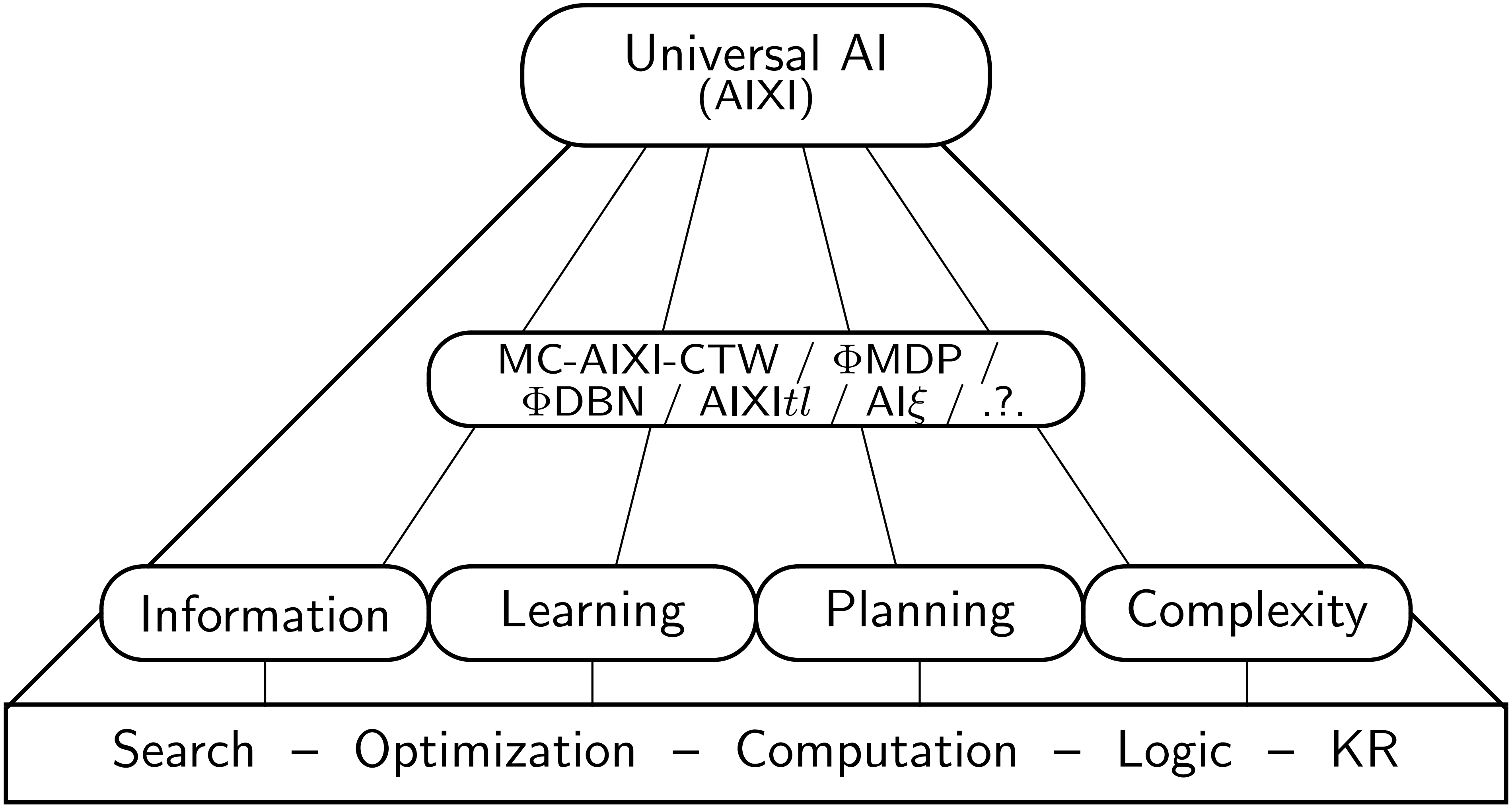
Feature Reinforcement Learning (ctd)

- Extensions:
  - Looping suffix trees for long-term memory [DSH12, DSH14a]
  - Structured/Factored MDPs (Dynamic Bayesian Networks) [Hut09a]

- Related:
  - Q-Learning for History-Based Reinforcement Learning [DSH13]
  - Convergence of Q-Learning Beyond MDPs [?]
  - Reinforcement Learning with Value Advice [DSH14b]
  - Extreme State Aggregation beyond MDPs [Hut14, Hut16, ?]

DISCUSSION

Intelligent Agents in Perspective



Agents = General Framework, Interface = Robots,Vision,Language

Aspects of Intelligence

are all(?) either directly included in AIXI or are emergent

Trait of Intell.	How included in AIXI
reasoning	to improve internal algorithms (emergent)
creativity	exploration bonus, randomization, ...
association	for co-compression of similar observations
generalization	for compression of regularities
pattern recognition	in perceptions for compression
problem solving	how to get more reward
memorization	storing historic perceptions
planning	searching the expectimax tree
achieving goals	by optimal sequential decisions
learning	Bayes-mixture and belief update
optimization	compression and expectimax
self-preservation	by coupling reward to robot components
vision	observation=camera image (emergent)
language	observation/action = audio-signal (emergent)
motor skills	action = movement (emergent)
classification	by compression
induction	Universal Bayesian posterior (Ockham's razor)
deduction	Correctness proofs in AIXItl

Miscellaneous Results

- Suitable versions of AIXI are limit-computable [LH15]
- How to discount future rewards [Hut06, LH14b]

Outlook

- Find optimality notions for generally intelligent agents which are strong enough to be convincing but weak enough to be satisfiable.
- More powerful and faster computational approximations of AIXI
- Social questions about AIXI or other Super-Intelligences: socialization, rewards, drugs, suicide, self-improvement, manipulation, attitude, curiosity, immortality, self-preservation.
- Training (sequence): To maximize informativeness of reward, one should provide a sequence of simple-to-complex tasks to solve, with the simpler ones helping in learning the more complex ones.

Outlook

- Address the many open theoretical questions in [Hut05].
- Bridge the gap between (Universal) AI theory and AI practice.
- Explore what role logical reasoning, knowledge representation, vision, language, etc. play in Universal AI.
- Determine the right discounting of future rewards.
- Develop the right nurturing environment for a learning agent.
- Consider embodied agents (e.g. internal$\leftrightarrow$external reward)
- Analyze AIXI in the multi-agent setting.

Thanks! Questions? Details:

[Hut05] M. Hutter. Universal Artificial Intelligence. Springer, Berlin, 2005. http://www.hutter1.net/ai/uaibook.htm

[VNH+11] J. Veness, K. S. Ng, M. Hutter, W. Uther, and D. Silver. A Monte Carlo AIXI approximation. *Journal of Artificial Intelligence Research*, 40:95–142, 2011.

[Leg08] S. Legg. Machine Super Intelligence. *PhD thesis*, IDSIA, Lugano, Switzerland, 2008.

[Hut12b] M. Hutter. One decade of universal artificial intelligence. In *Theoretical Foundations of Artificial General Intelligence*, pages 67–88. Atlantis Press, 2012.

[Lat14] T. Lattimore. Theory of General Reinforcement Learning. *PhD thesis*, Research School of Computer Science, Australian National University, 2014.