XAI and Strategy Extraction via Reward Redistribution

Sepp Hochreiter

JKU Institute for Machine Learning
JKU LIT Artificial Intelligence Lab
ellis LINZ unit
IARAI the institute of advanced research in artificial intelligence

XAI Goals

- Build up trust:
  - explain the machine's decisions
  - make the machine's decisions comprehensible
- Verify and certificate:
  - verify decisions
  - certificate procedures and evaluations
  - robustness (generalization to new situations)
  - safety
- Avoid biases:
  - input data (ethnic groups, gender, situations)
  - output data / teacher / target → human bias, human errors

XAI Goals

Why did an algorithm recognize an object?

"Clever Hans" predicts right for the "wrong" reason: recognizing

- boats by the presence of water
- trains by the presence of rails
- horses by the presence of a copyright watermark
- "Husky" (not "Wolf") by the presence of snow
- table tennis ball by the presence of a table tennis table
- basket ball by the presence of an indoor sports floor

Generalization is questionable

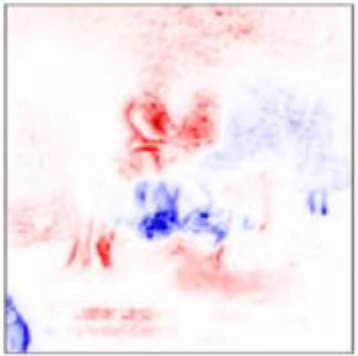
XAI Goals

"Clever Hans" effect: horses recognized by a copyright watermark

Original Image

Standard LRP

Original Image



Standard LRP

XAI: Correct / Helpful

Is the explanation **correct**?

- Explaining procedure gives the same model result
- Explaining procedure can substitute the model
- Explaining procedure leads to the same policies / returns

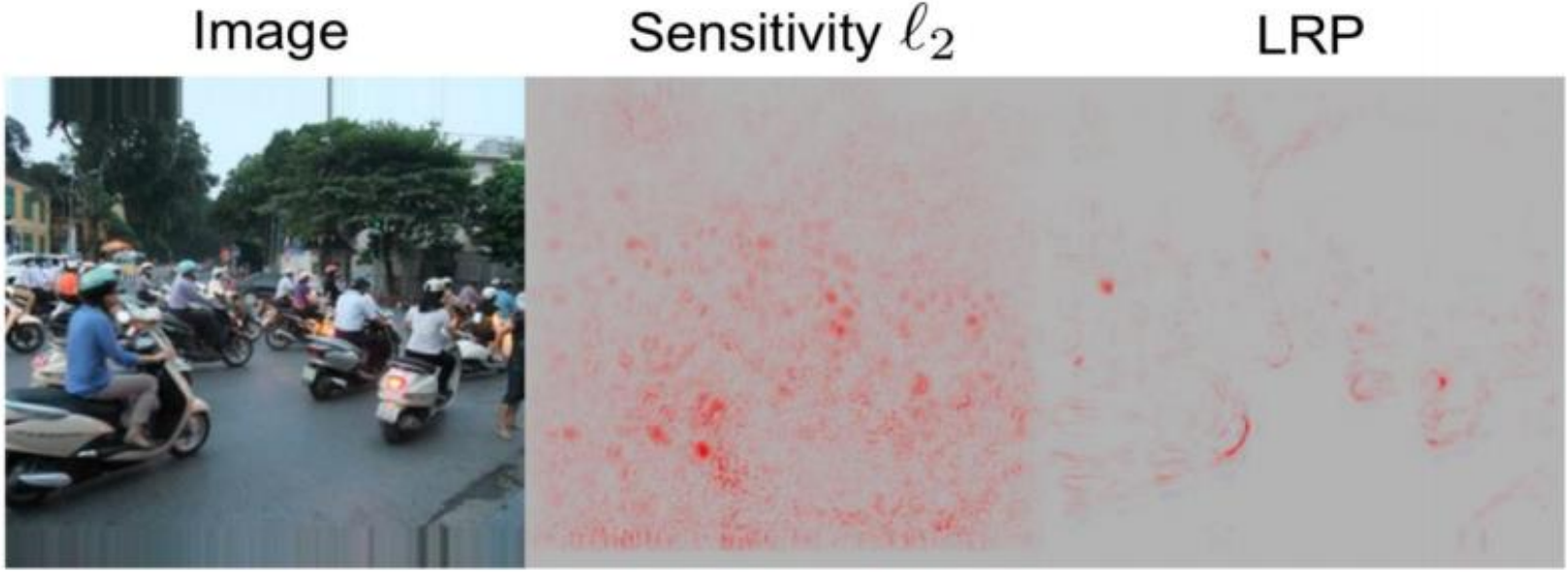
Does the explanation **help**?

- Better understandable than the original model
- Uses human concepts
- Less complex (linear, few connections, few nodes)
- Less dependencies: no side effects, no affects on future states

XAI Methods

1. Simple surrogate functions to explain the predictions
2. Testing the response of the model
   - Sensitivity analysis → gradient-based (BP through a model)
   - Occlusion of inputs (masking out inputs regions)
   - Maximal response inputs (inputs that maximize an output)
3. Contribution analysis uses or analyzes the model on examples
   - Layer-wise relevance propagation (LRP)
   - Integrated gradient (IG)
   - Difference of predictions
4. Meta-explanation of model behavior
   - Analyze learned representations

XAI: Sensitivity Analysis



Sensitivity Analysis:

→ *"what makes this image less / more 'scooter' ?"*

LRP / Taylor Decomposition:

→ *"what makes this image 'scooter' at all ?"*

XAI: Contribution Analysis

Contribution analysis: our XAI focus

Contribution gives immediate feedback: easy learning, understandable

Contribution adjusts the expectation of the outcome

XAI: Credit Assignment

Contribution Analysis = Credit Assignment

Contribution analysis: analyze a machine learning model with respect to the contributions of inputs to the output.

1. **XAI:** to explain a model

2. **Reinforcement learning:** to learn from a model

Credit Assignment

Assigning credit for a received reward to previously performed actions is one of the central tasks in reinforcement learning.

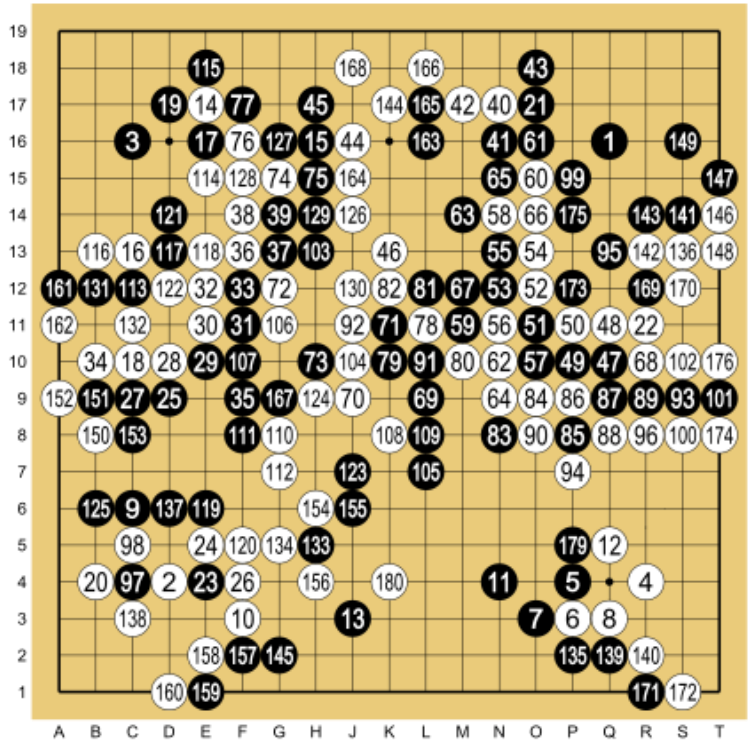
One of the great challenges is long-term credit assignment:

- delayed rewards
- sparse rewards
- episodic rewards

Episodic rewards:

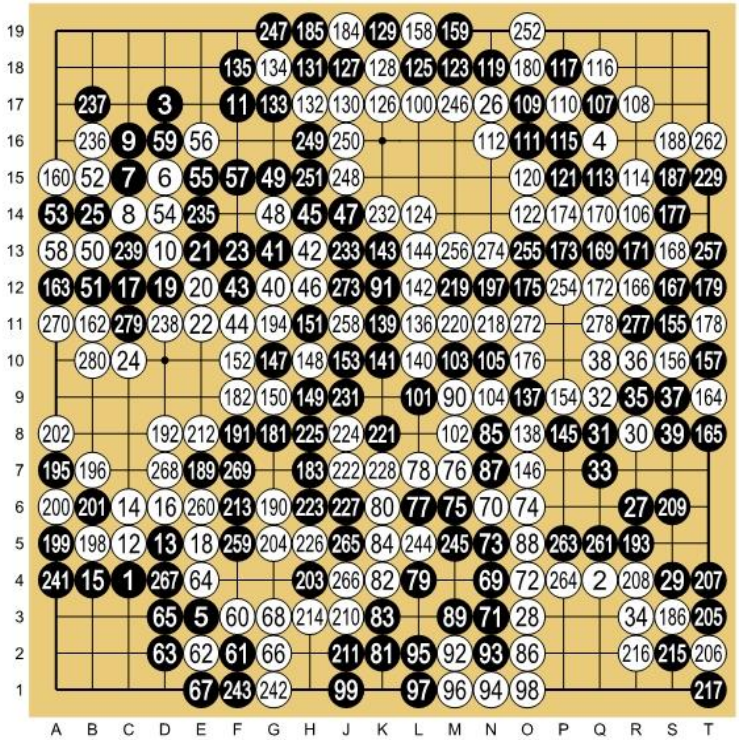
- Achieving a goal
- Completing a task
- Accomplishing something

Credit Assignment



Lee Sedol (W) vs AlphaGo (B) - Game 4

177 at 51 178 at 57



Lee Sedol (B) vs AlphaGo (W) - Game 5

118 at 107 161 at 25 230 at 148 234 at 53 240 at 200 253 at 184

271 at 25 275 at 168 276 at 151

Reinforcement Learning

Model-free reinforcement learning with **strategic decisions**:

- logistics
- drug design
- energy
- self-driving cars
- optimization of traffic and smart cities (air pollution)
- environment and climate change


DOTA 2
OpenAI


StarCraft II
WINGS OF LIBERTY
StarCraft II
DeepMind
JKU

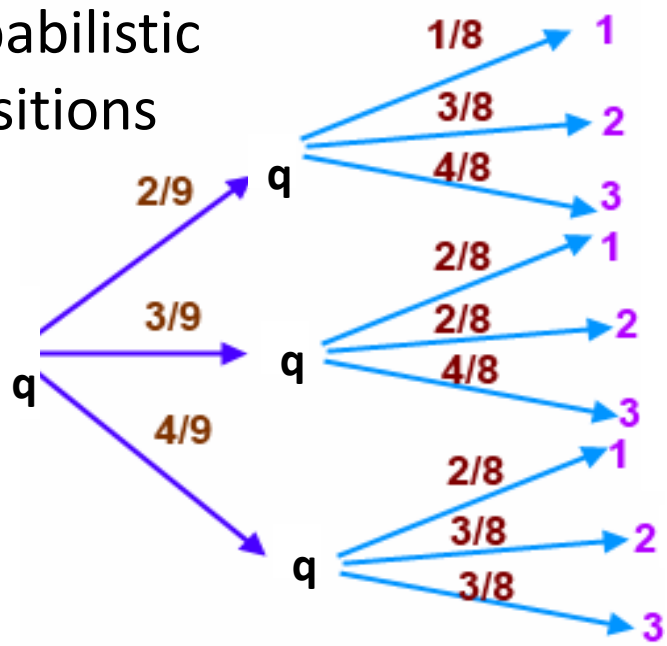
Delayed Rewards

Strategic decisions lead to **delayed rewards**:

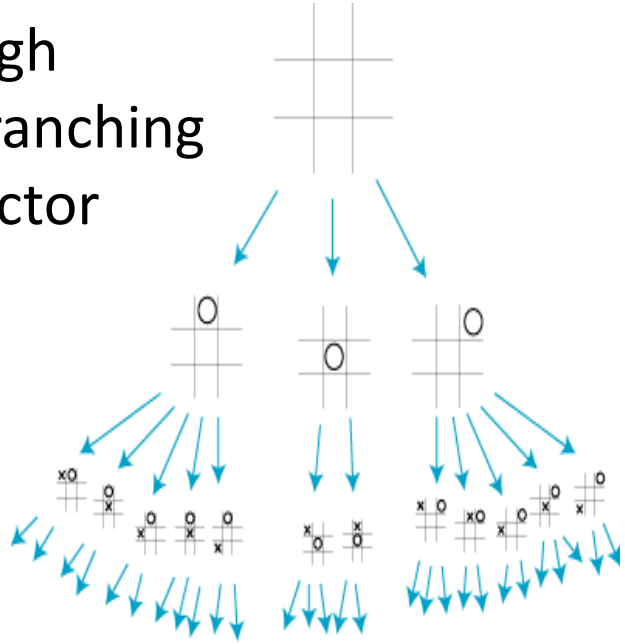
- actions cause reward or penalty that is obtained much later
- distracting rewards may be present
- credit assignment problem: what action was responsible

Delayed Rewards: Problem

probabilistic transitions



high branching factor



- averaging over many possible futures (Monte Carlo)
- propagating reward back has exponential decay (temporal difference)

Our Goal

All future expected reward is zero: it is given immediately

- reward is the **change in the expected return**
  - increase of expected return → positive reward
  - decrease of expected return → negative reward
- immediately adjust the return expectation

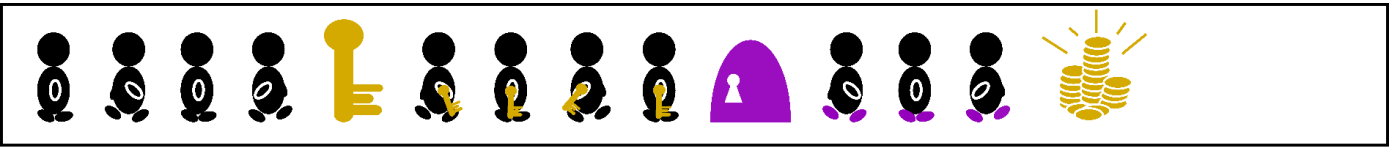
Reward Redistribution

- complex tasks: hierarchical with sub-tasks or sub-goals
- **value function is step function**: change in return expectation

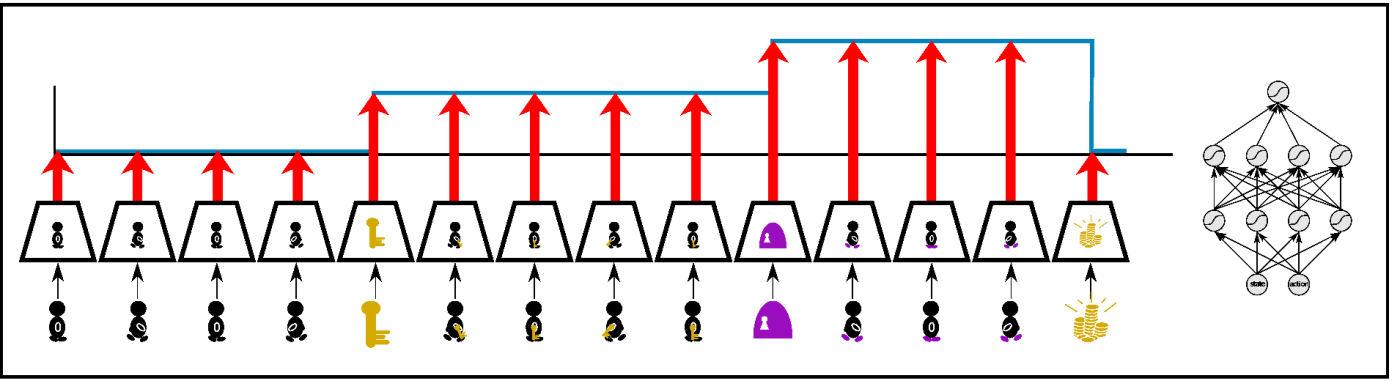
example

- getting key → increases the probability of obtaining the treasure
- opening door → increases the probability of obtaining the treasure

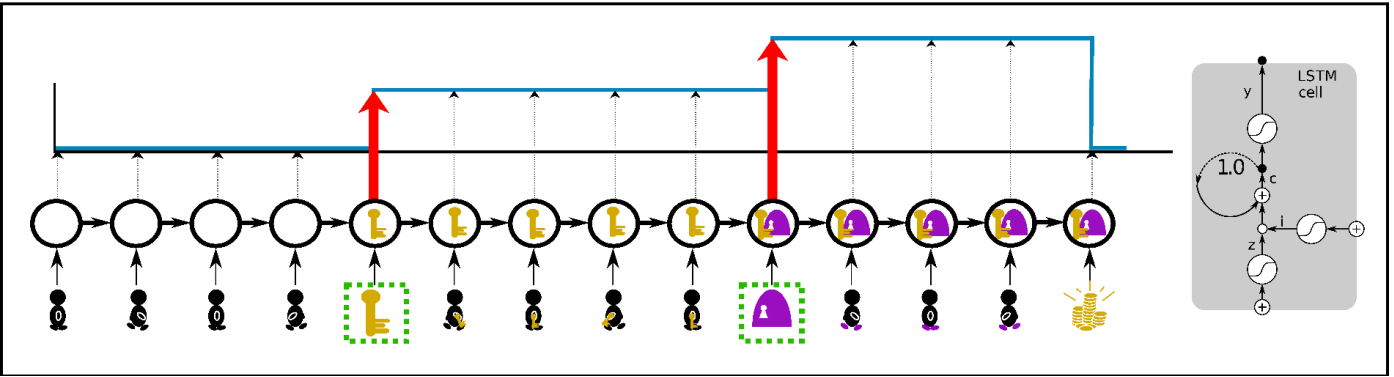
Reward Redistribution



Learning step functions



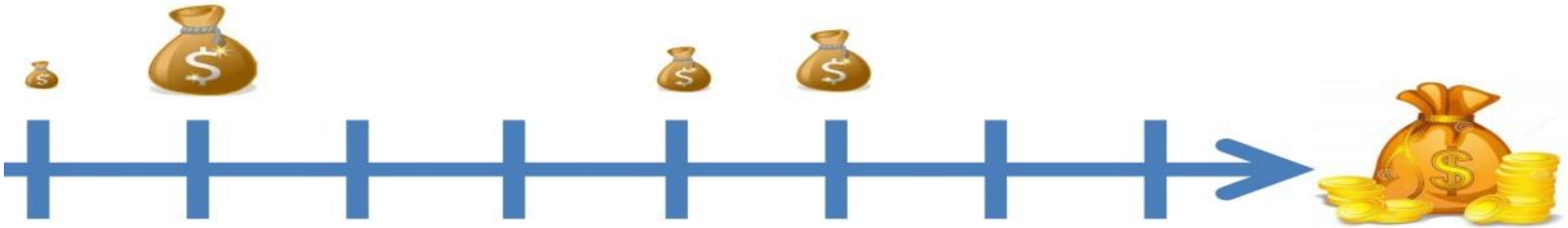
fully connected networks
→ every state-action



LSTM or alignment
→ memorizing steps
→ much more efficient

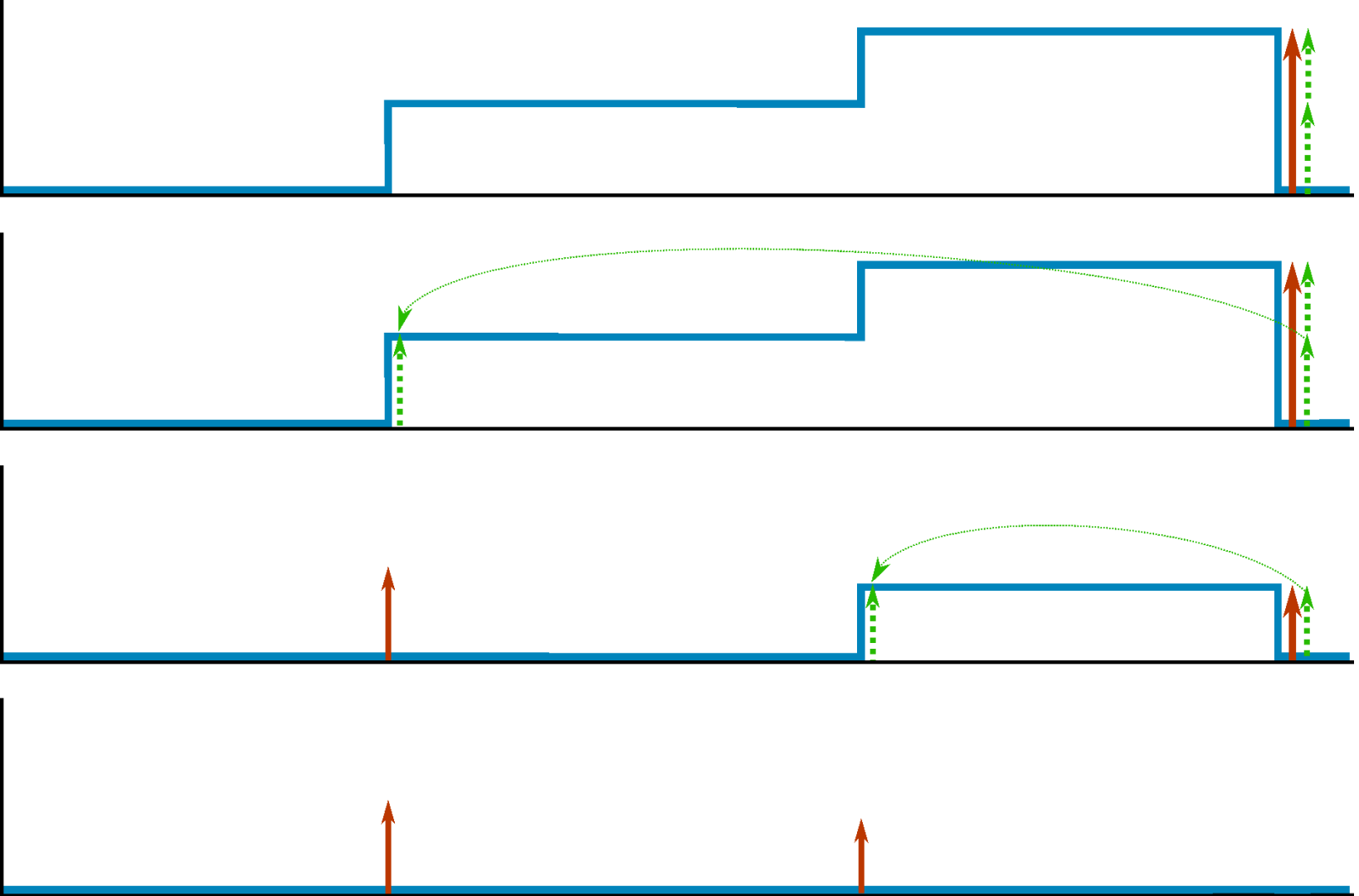
Reward Redistribution

reward redistribution: give reward when return expectation changes



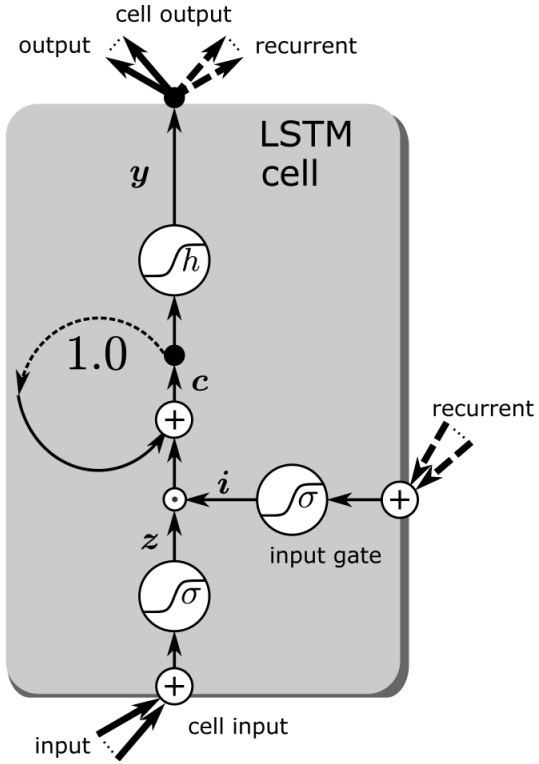
reward redistributions do not change optimal policies.

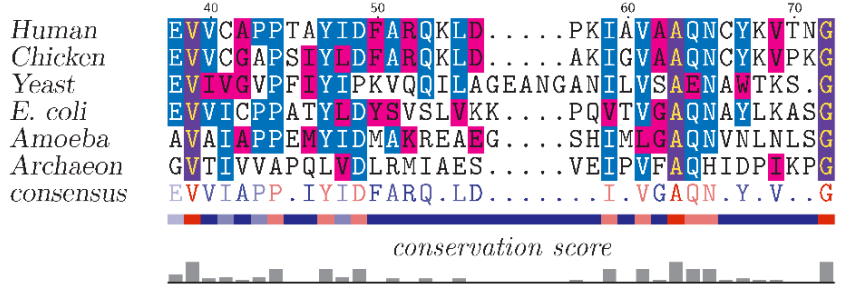
GOAL: all future expected reward is zero since already given

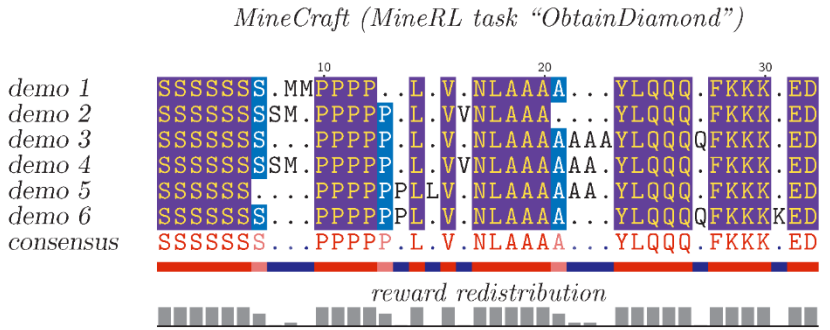


Reward Redistribution

Steps are identified by LSTM or by alignment model.



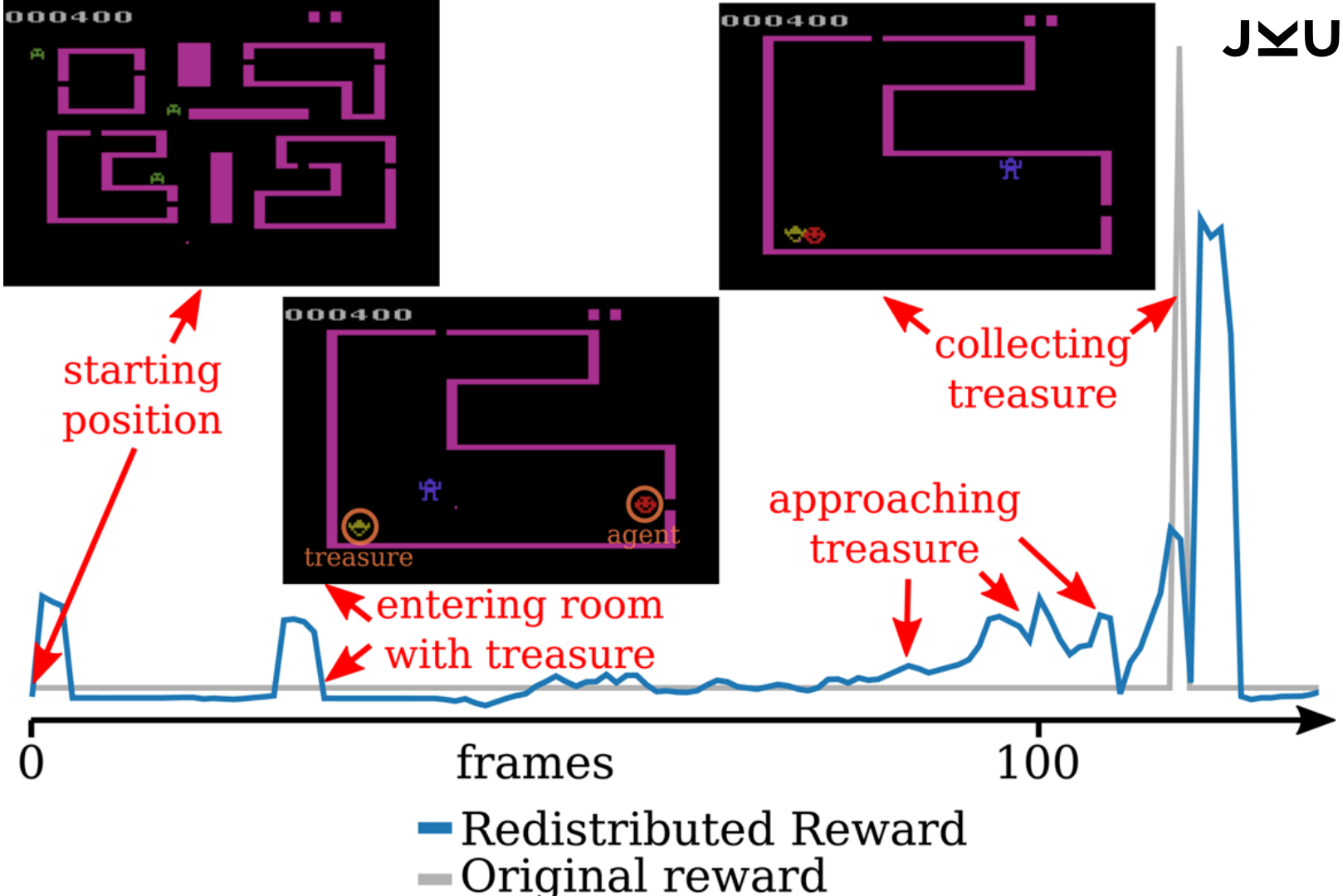




XAI: Reward Redistribution

Reward redistribution

- explains the prediction (what inputs contributed to the prediction)
- explains the consequences of actions (what happens in the future)
- explains a policy or a strategy (why is an agent better than another)
- explains the performance of an agent (why did it achieve the goal)


000400
000400
000400
starting position
collecting treasure
treasure
agent
approaching treasure
entering room with treasure
0
frames
100
Redistributed Reward
Original reward

END