Adaptive Modular Architectures for Rich Motor Skills

SEVENTH FRAMEWORK PROGRAMME

Research Institute for Cognition and Robotics

Universität Bielefeld

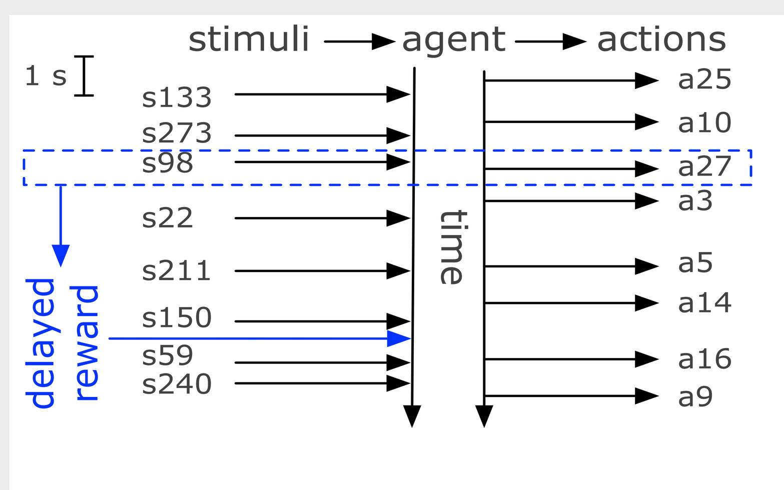
Short and long term plasticity as cause-effect hypothesis testing in robotic ambiguous scenarios

Andrea Soltoggio

Neurorobotic systems learn with difficulties when a continuous flow of information and delays make the cause-effect relationships ambiguous. The Hypothesis Testing (HT) plasticity proposed in this study models learning dynamics that account for ambiguity in the sensory-motor information flow improving drastically discrimination capabilities and memory capacity with respect to previous models. The new rule models consolidation to long-term memory and helps solve the plasticity-stability dilemma.

Ambiguous cause-effect relationships with delayed rewards

What sequence of stimuli and actions effectively causes a reward? One single instance cannot disambiguate the relationship if delays and asynchronous information flow are present.



Rarely Correlating Hebbian Plasticity (RCHP) (equivalent to R-STDP)

$$\mathrm{RCHP}_{ji}(t) = \begin{cases} +1 & \text{if } v_j(t - t_{pt}) \cdot v_i(t) > \theta_{hi} \\ -1 & \text{if } v_j(t - t_{pt}) \cdot v_i(t) < \theta_{lo} \\ 0 & \text{otherwise} \end{cases}$$

eligibility traces $\quad c_{ji}(t + \Delta t) = c_{ji}(t) \cdot e^{\frac{-\Delta t}{\tau_c}} + \mathrm{RCHP}_{ji}(t) \quad .$

modulation $\quad d(t + \Delta t) = d(t) \cdot e^{\frac{-\Delta t}{\tau_m}} + \lambda\, r(t) + b \quad .$

weight update $\quad \Delta w_{ji}(t) = c_{ji}(t) \cdot d(t)$

Hypothesis Testing Plasticity. Addition of

1) short-term and long-term components

2) hypothesis testing potentiation and depression with negative baseline modulation (term b)

$$\mathrm{RCHP}^{+}_{ji}(t) = \begin{cases} +1 & \text{if } v_j(t - t_{pt}) \cdot v_i(t) > \theta_{hi} \\ 0 & \text{otherwise} \end{cases}$$

eligibility traces $\quad c_{ji}(t + \Delta t) = c_{ji}(t) \cdot e^{\frac{-\Delta t}{\tau_c}} + \mathrm{RCHP}^{+}_{ji}(t) \quad .$

short-term weight dynamics $\quad \dot{w}^{st}_{ji}(t) = -w^{st}_{ji}/\tau_{st} + m(t) \cdot c_{ji}(t)$

Consolidation of hypotheses (weights)

$$W_{ji}(t) = w^{st}_{ji}(t) + w^{lt}_{ji}(t)$$

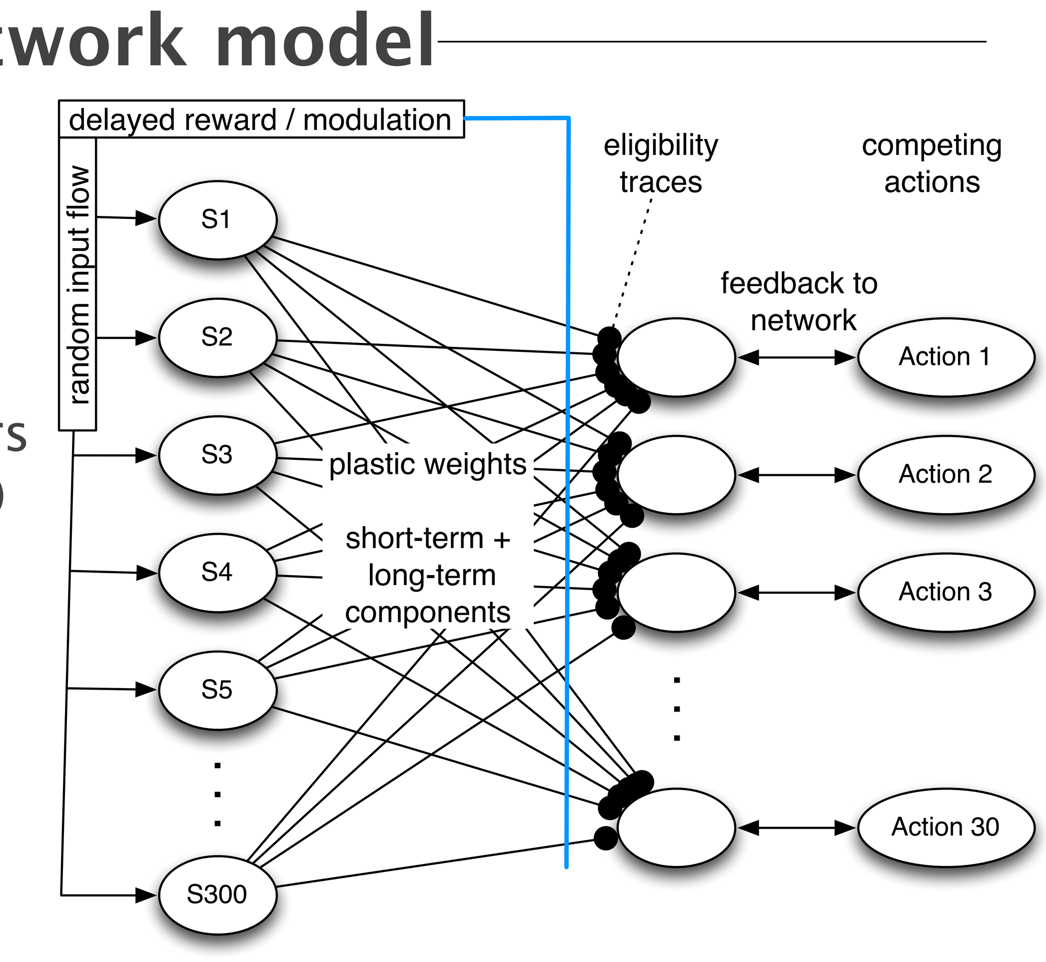
$$\dot{w}^{lt}_{ji}(t) = \rho \cdot H(w^{st}_{ji}(t) - \Psi)$$

HT-plasticity in a network model

Problem
- large input-output flow
- delayed rewards
- ambiguous cause-effect relationships
- no external reward-predictors (i.e. expected average rewards)

Network
- 300 inputs, 30 outputs
- 9000 stimulus-action pairs

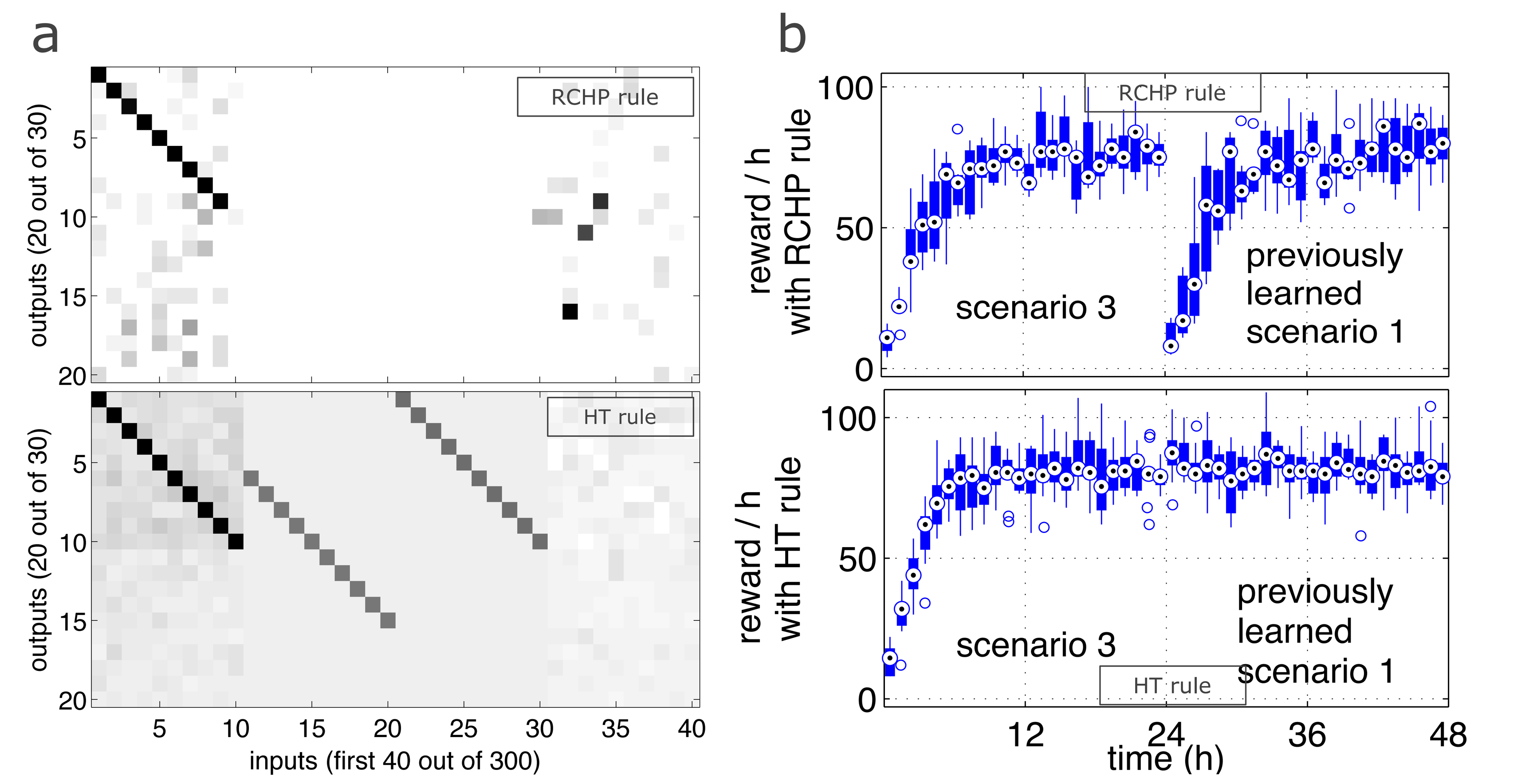
Advantages
- short-term components represent hypotheses
- long-term components represent established facts
- HT allows for learning over multiple scenarios
- HT plasticity allows for higher discrimination of true cause-effect relationships



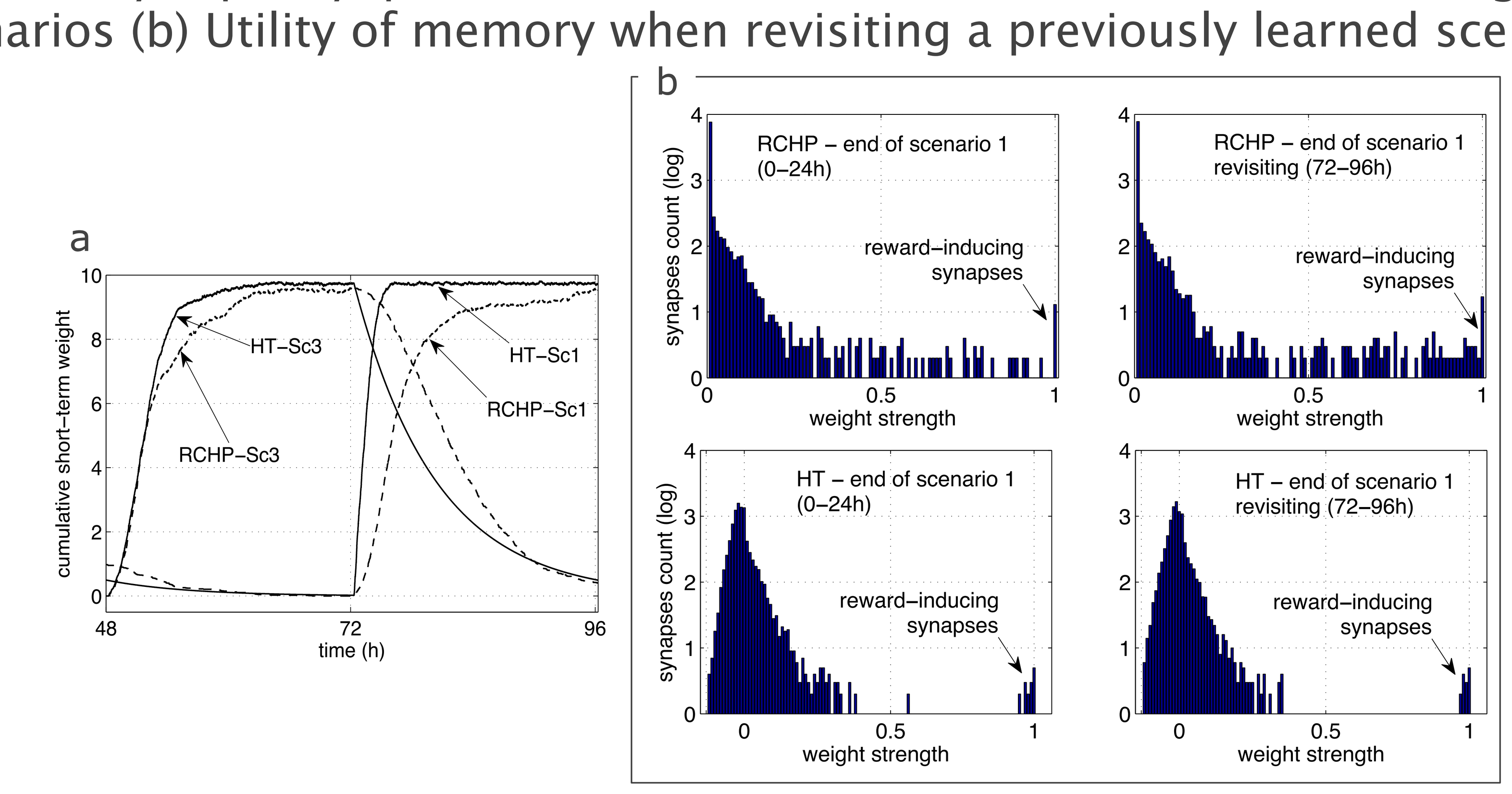
Suitable for
- using memory capacity of network
- modeling knowledge-discovering in ambiguous environments
- studying classical and operant conditioning over long time scale and multiple associations

References.
- Soltoggio, Reinhart, Lemme, Steil. Learning the rules of a game: neural conditioning in human-robot interaction with delayed rewards. ICDL 2013
- Soltoggio, Lemme, Reinhart, Steil. Rare neural correlations implement robotic conditioning with delayed rewards and disturbances. Frontiers in Neurorobotics (2013)

SIMULATIONS



(a) Memory capacity, preservation of information over different learning scenarios (b) Utility of memory when revisiting a previously learned scenario

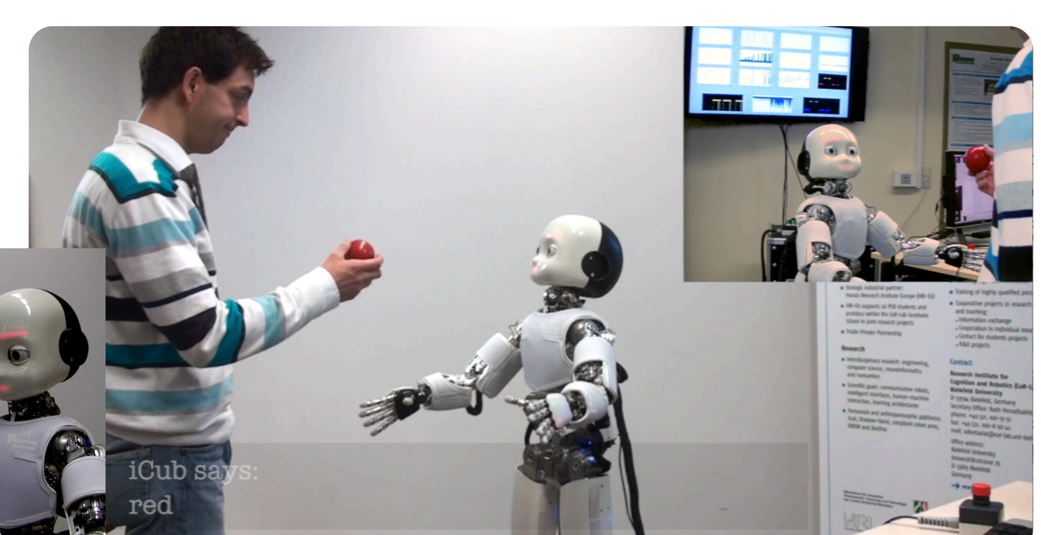


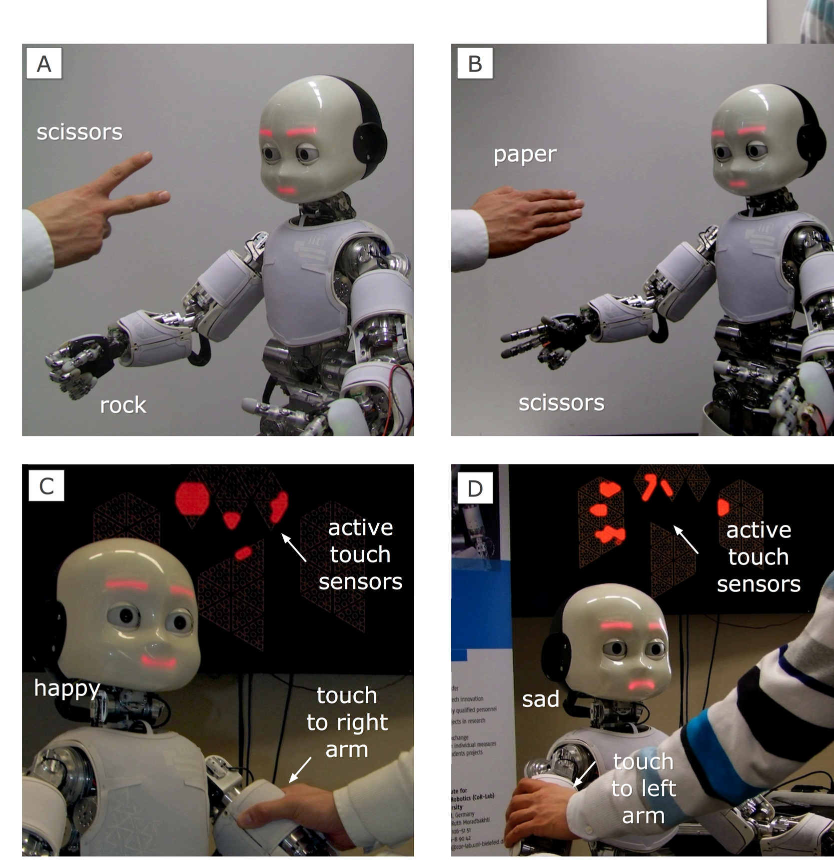
(a) Improved exploration speed (b) Discrimination capability and robustness of the HT rule with respect to the previous approaches

NEURO-ROBOTICS APPLICATIONS : combination of skills

Previous experiments with RCHP can now be all combined and expanded thanks to the increased memory capabilities and different learning scenarios

iCub plays Rock-Paper-Scissors



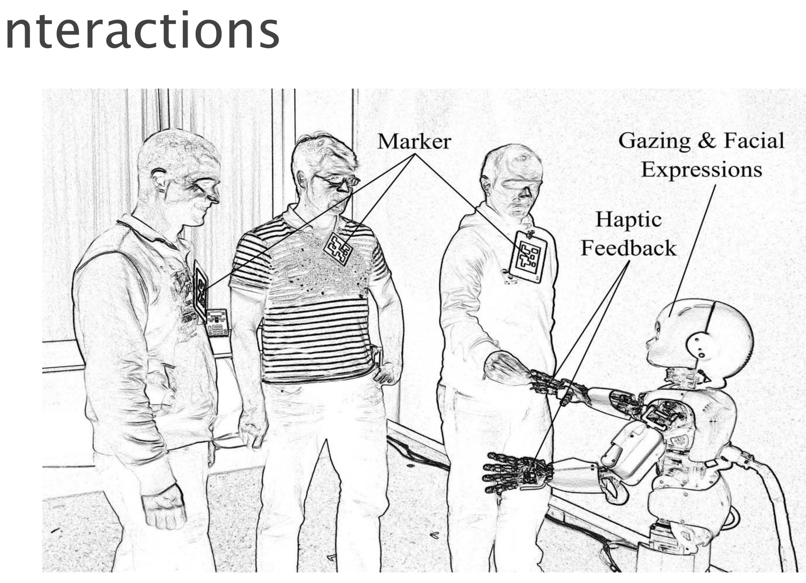


Uncertain feedback signals in human-robot interaction

Classical conditioning: iCub makes new friends

iCub learns the colors

Operant conditioning with reversal learning. Learning from trial and error in human-robot interactions



Conclusion. The HT-plasticity improves both exploration and exploitation capabilities of the network. It increases memory capabilities by preserving established learned relationships, it detects coincidental facts as irrelevant and allows for the combination of more learning scenarios and skills.