Optimizing Multi-Context Motor Adaptation using Model-Free Reinforcement Learning

Russell Jeter (2), **Dmitrii Todorov** (1), Yaroslav Molkov (2); (1) = Laboratoire d'Imagerie Biomedicale / INSERM / CNRS, Paris, France; (2) = Georgia State University, Atlanta, GA, USA

Introduction

Motor rehabilitation (e.g. after stroke) relies on transfer from trained tasks to real-world contexts. Usually it involves a series of tasks that are performed many times (many trials).

Current protocols show limited generalization from rehab training to real life. Hypothesis: this happens because rehab motor task schedules are heuristic (random or blocked).

Goal: Automate curriculum design to maximize structured learning.

Scientific question: Given history of participant behavior up to *trial N* what context to present to a participant on *trial N+1* so that the resulting performance is optimal (or at least, better than a simplest reasonable benchmark)?

Methods: train a simple deep reinforcement learning (RL) on synthetic data generated by an advanced motor adaptation model reproducing human visuomotor adaptation experiment

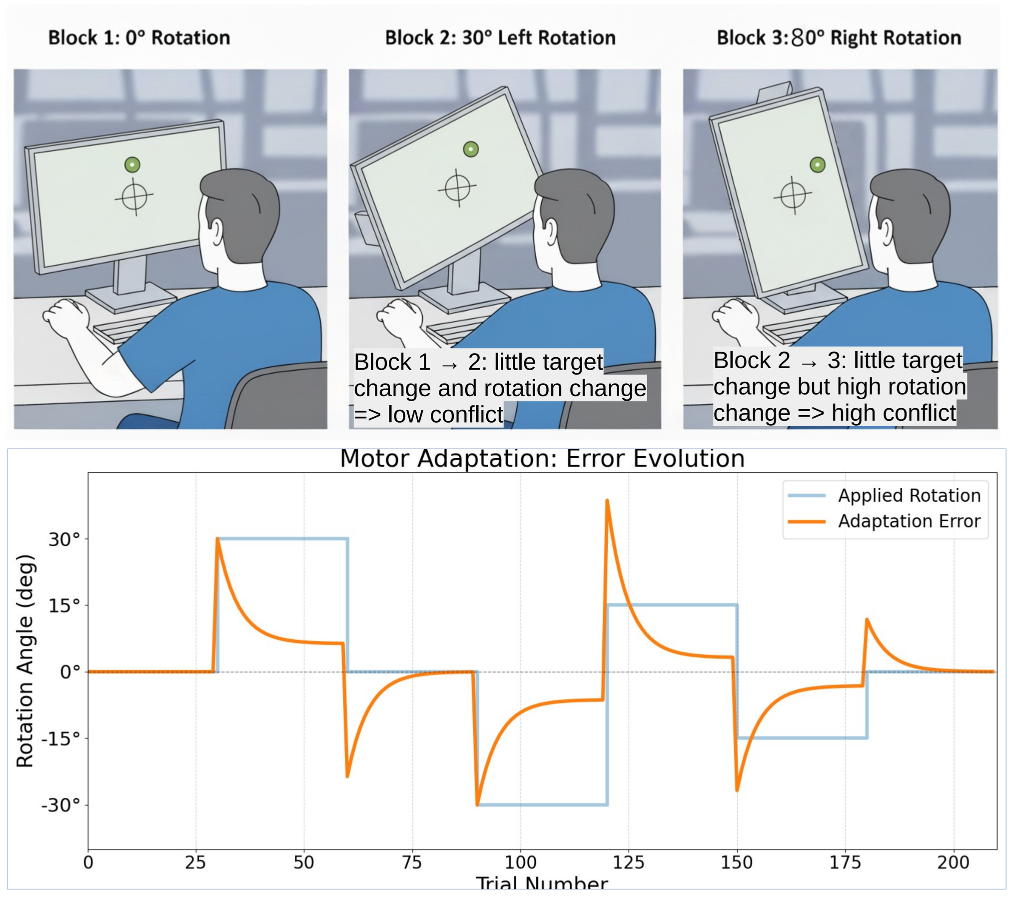


Fig1: Schematic of a blocked motor adaptation experiment. **Top**: (conceptual) illustration of blocked adaptation to three different visuomotor rotations with three different targets. **Bottom**: behavior of a simple state space model in the blocked motor adaptation experiment (with perfect error generalization and fixed adaptation rate, unlike in further simulations)

Adaptation model: $N_{contexts}$ - dimensional state space model with generalization and adaptive error-dependent learning rate

$$e_{t+1} = \text{perturbation}_{t+1} - x_{t+1}^{c_{t+1}} + \eta_{t+1}, \quad \eta_{t+1} \in U(-0.5, 0.5)$$

$$w_{t+1}^{(b)} = w_t^{(b)} + \Lambda \cdot \text{sign}(e_t, e_{t+1}) \cdot \text{gauss basis}^{(b)}(e_t), \quad 1 \le b \le N_{contexts}$$

$$\lambda_{t+1} = \sum_b w_{t+1}^{(b)} \cdot \text{gauss basis}^{(b)}(e_{t+1})$$

$$x_{t+1}^i = R \cdot x_t^i + S_{i,c_{t+1}} \lambda_{t+1} e_{t+1}, \quad 1 \le i \le N_{contexts}$$

where R is the retention rate, $S_{i,j}$ is the similarity score (between 0 and 1) of contexts i and j (see [2]), λ_t is the error sensitivity (adaptive learning rate) evaluated after trial t as in Memory of Errors[1] model, meaning that it increases when two preceding trials' errors had the same sign, and decreases otherwise, Λ is the learning rate for the adaptation rate itself. Error space is covered by 100 Gaussian basis functions

Contexts

What is a *context*?
Each context is a pair of (target, perturbation)
We use 10 context in total

Context distribution
Perturbations are drawn from a fixed distribution before the simulation is started
Perturbations are independent from the targets
Perturbation distribution can be either **binary {-1,1}** or **uniform [-1,1]**

Contexts delivery are delivered
* either randomly using a **Markov chain** rule on $N_{contexts}$ states
* or following a trained RL policy

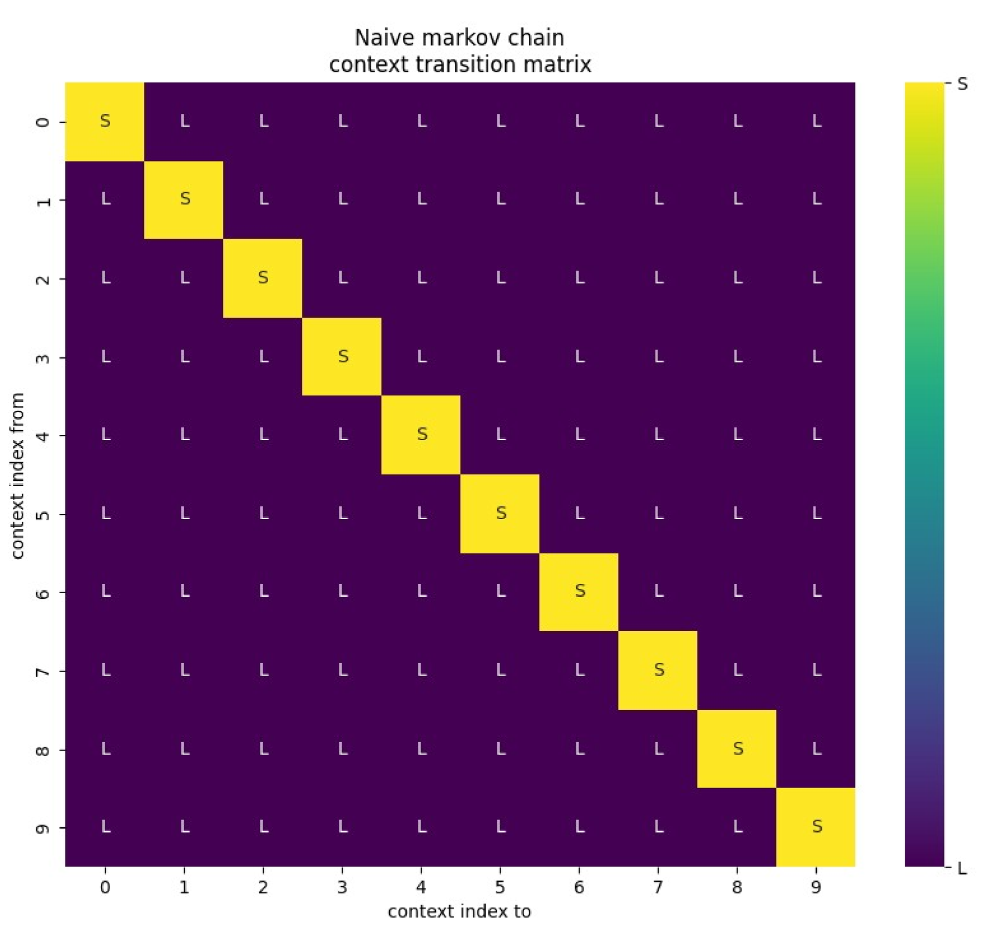
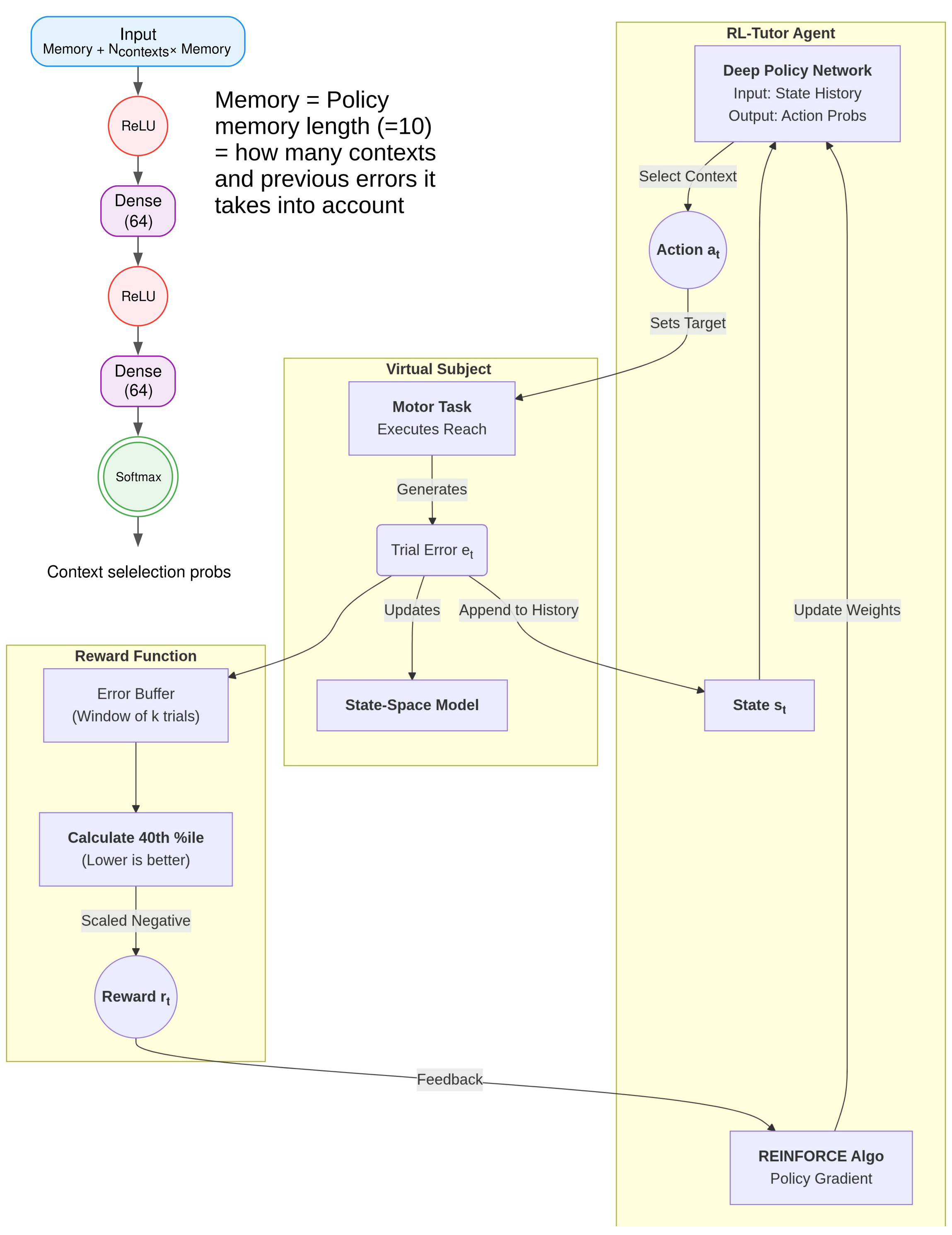


Fig 2: Markov chain transition probability matrix. Stay with probability S and leave to one of the other contexts with probability $L{=}(1{-}S)/(N{-}1)$

The Closed-Loop RL Tutor Framework

Train RL policy using REINFORCE algorithm



Quantile reward:
1. Compute errors in all contexts at the current step
2. Take 40%-th

Conceptual: comparing how **best** contexts perform, don't punish too hard contexts

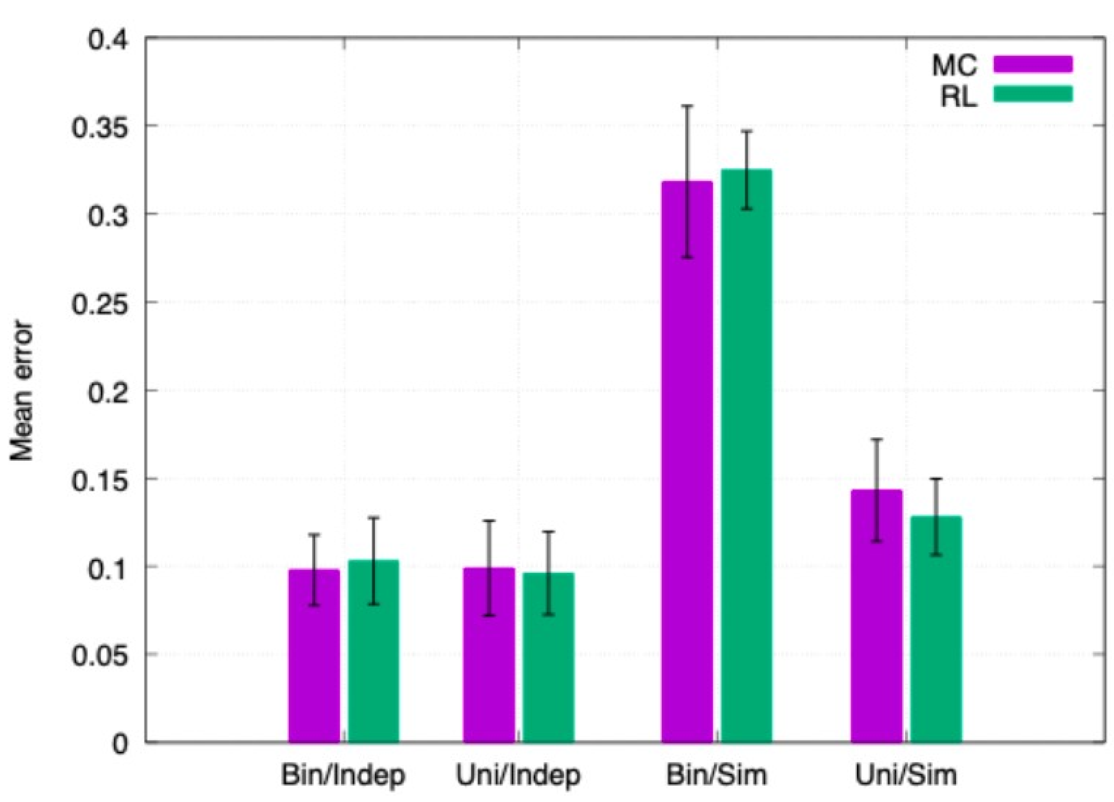


Fig 4. Performance error comparison for simulated motor learning with Markov chain of length 2 vs. RL-Tutor curricula. Performance error measures the **mean absolute error** across all contexts. The RL-Tutor (red) consistently outperforms the Markov-2 curriculum, where the next context is determined by a Markov chain with a probability of self-transition of 0.5, (blue) across all four tested schemes. The schemes combine perturbation types (Bin: binary ±1; Uni: uniform [-1, 1]) with context similarity types (Indep: independent; Sim: locally similar).

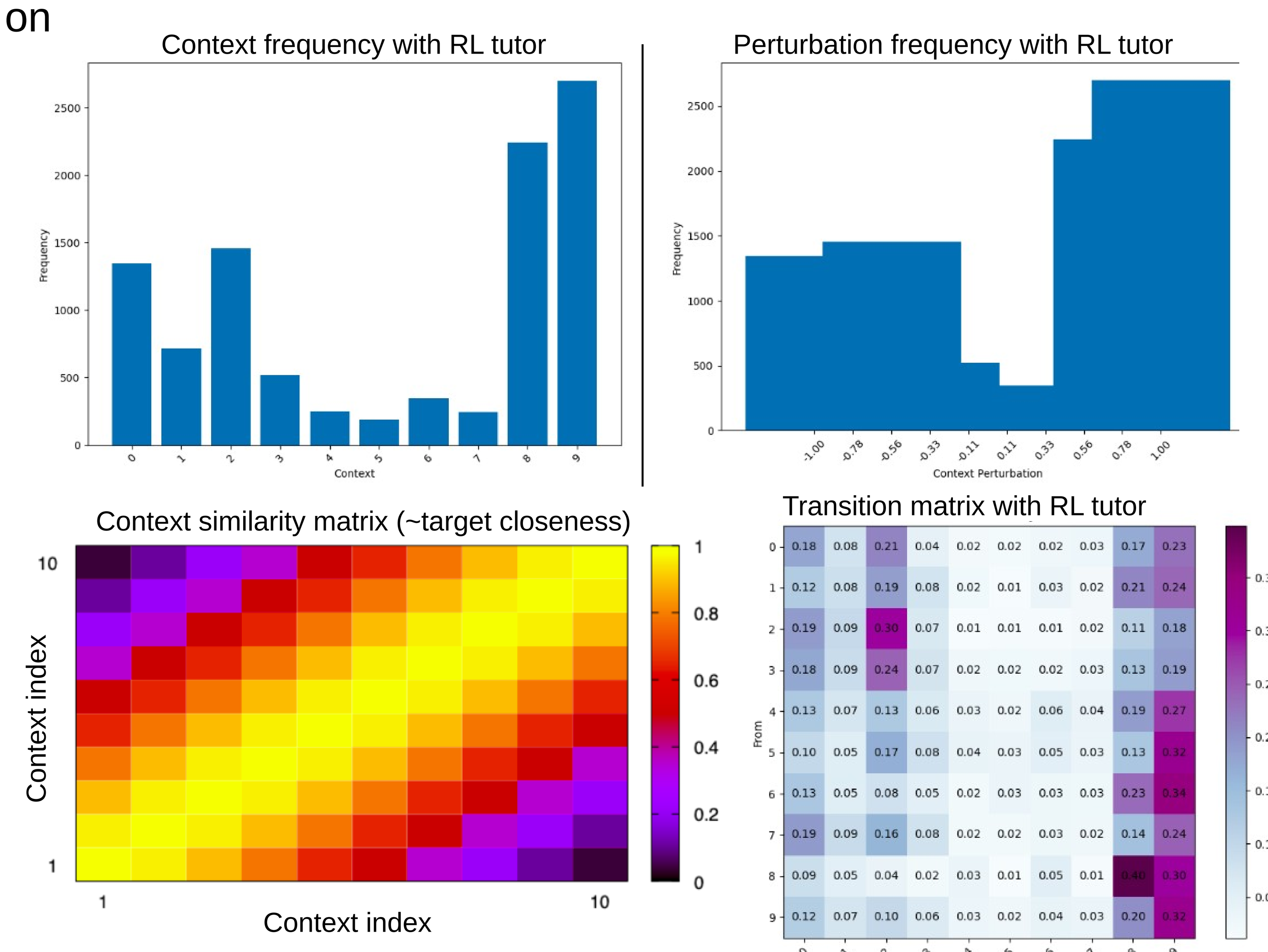


Fig 3: Context similarity matrix used for simulation (smooth decay of generalization) and results of RL tutor simulation

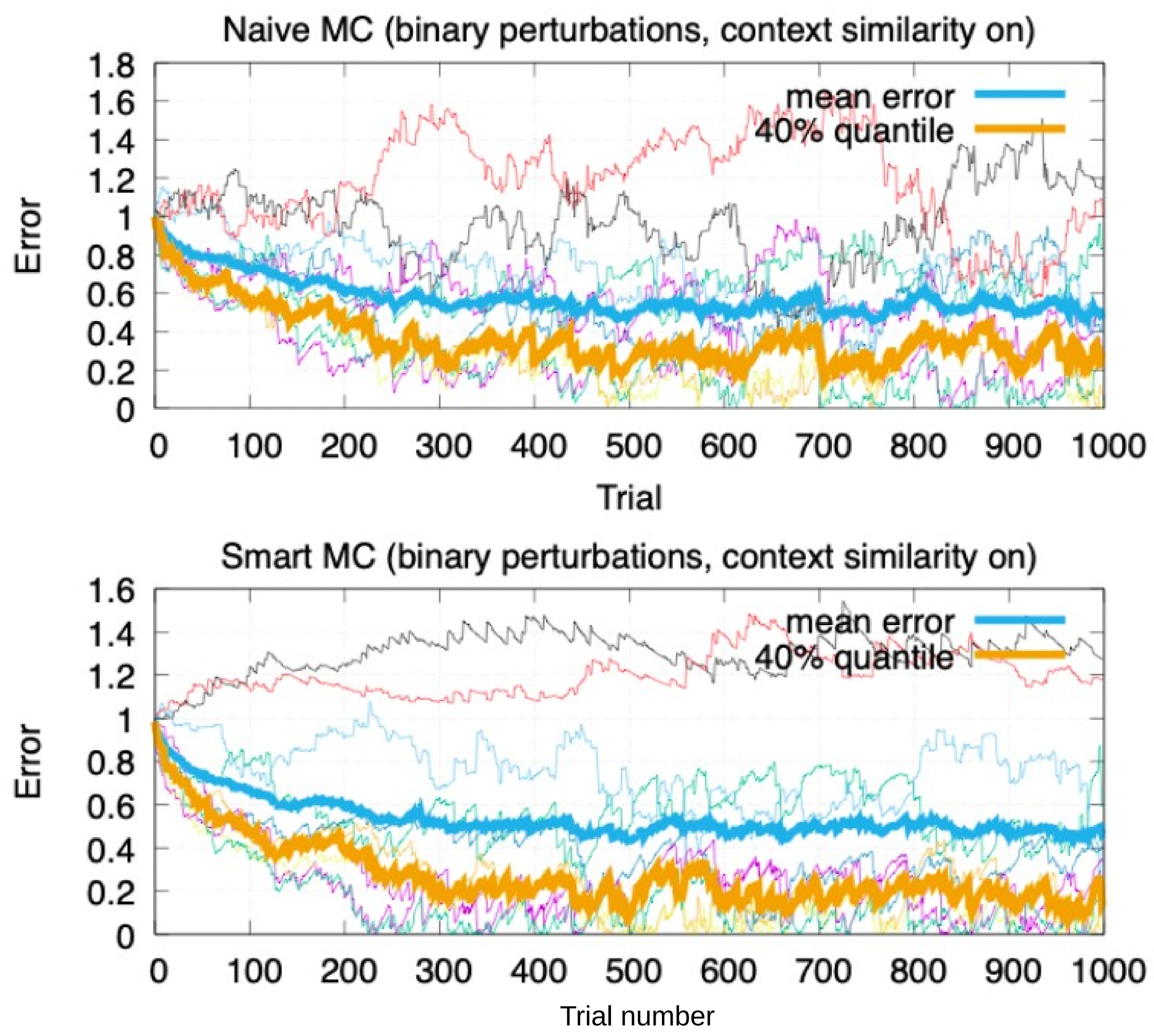


Fig. 5: Simulation of the virtual participant model with contexts delivered by either naive Markov Chain or by an RL policy. Different traces are correspond to error traces in different contexts (including not-presented ones). All angles (and thus errors) scaled to -1,1 range

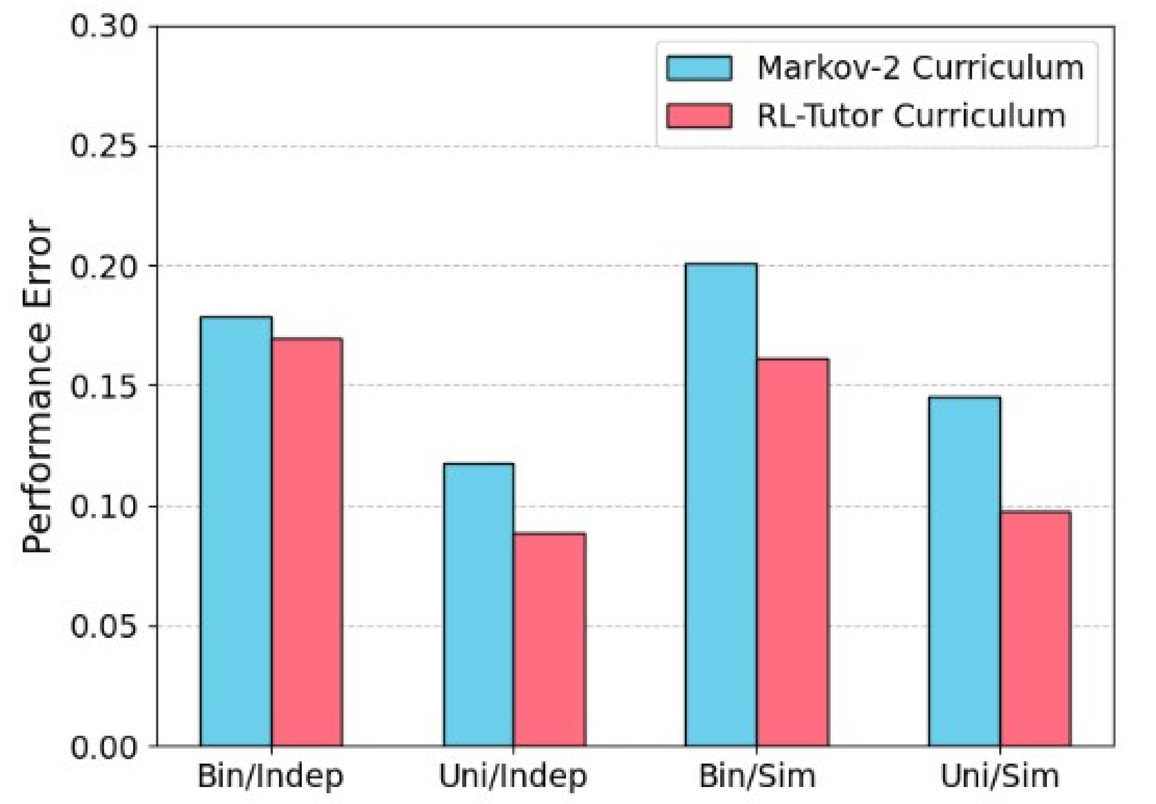


Fig 6. Same as Fig. 4 but for the **lower 40th percentile** absolute error across all contexts.

Results

- Mean (and median) are not very different between naive and and advanced context switching, **but lower percentiles are**
- When contexts are independent all schedules perform similarly
- When context similarity is taken into account, RL-guided adaptation is more efficient than the naive block switching for perturbations uniformly at random associated with targets

Significance

- Neuro: Demonstrates that structured learning can be actively managed by algorithmic tutors.
- AI: Model-free RL can solve curriculum design problems with complex, non-standard objectives (quantile loss).
- Clinical: A step toward personalized, auto-adaptive stroke rehabilitation protocols that prioritize skill consistency.

Future directions

- Apply to real data
- Test in an actual experiment
- Supply RL tutor with neural data as well

References

[1]Herzfeld DJ, .., Shadmehr R. A memory of errors in sensorimotor learning. Science. 2014

[2]Donchin O, …, Shadmehr R. Quantifying Generalization from Trial-by-Trial Behavior of Adaptive Systems that Learn with Basis Functions: Theory and Experiments in Human Motor Control. J Neurosci. 2003

[3]Williams RJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach Learn. 1992