Adaptive Interaction of Persistent Robots to User Temporal Preferences

Kim Baraka[†], Manuela Veloso[††]

[†] kbaraka@andrew.cmu.edu, Robotics Institute, Carnegie Mellon University

[††] mmv@cs.cmu.edu, Computer Science Department, Carnegie Mellon University

International Conference on Social Robotics (ICSR), Paris, October 2015

Abstract. We look at the problem of enabling a mobile service robot to autonomously adapt to user preferences over repeated interactions in a long-term time frame, where the user provides feedback on every interaction in the form of a rating. We assume that the robot has a discrete and finite set of interaction options from which it has to choose one at every encounter with a given user. We first present three models of users which span the spectrum of possible preference profiles and their dynamics, incorporating aspects such as boredom and taste for change or surprise. Second, given the model to which the user belongs to, we present a learning algorithm which is able to successfully learn the model parameters. We show the applicability of our framework to personalizing light animations on our mobile service robot, CoBot.

MOTIVATION

Persistent robots call for a new paradigm for designing **adaptive interaction** with humans.

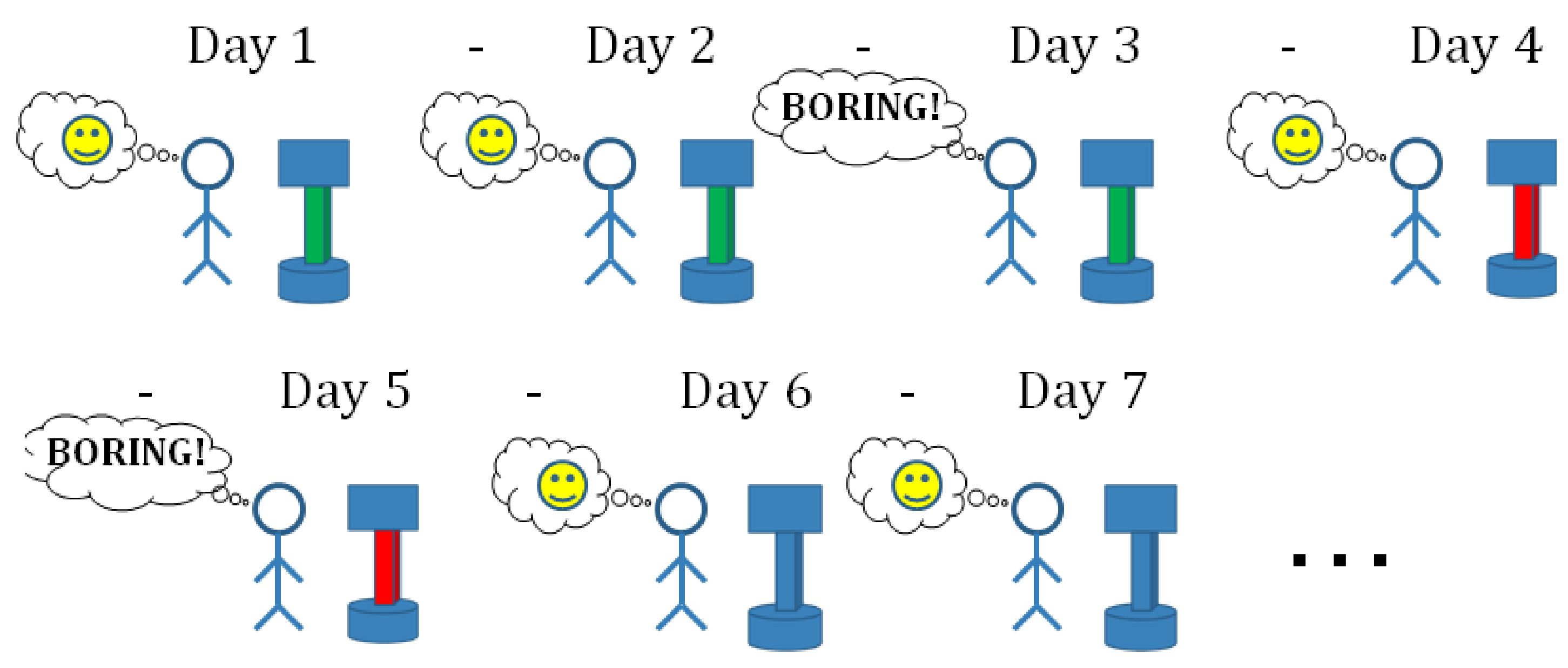


Fig. 1: Example where the robot keeps the user entertained by changing its appearance intelligently

Goal: Enable a robot to select **actions** so as to achieve **personalized persistent interaction** by **learning the dynamics** of user **preferences** over prolonged periods of time.

How? Learning through rewards over **time** and appropriate action selection

Learning sets → **Learning sequences**

MAIN CONTRIBUTIONS

- Three **models** for **dynamic** daily user **preferences**, accounting for **surprise**, **fatigue** and taste for **change**
- A **setup** to **adapt** to these models
- Robust **algorithms** to **learn** online the model parameters from user rewards

PROBLEM FORMULATION

- **Time steps** (*i*) ≡ encounters between robot and user
- **Actions** a_i ≡ interaction options
- **Rewards** $r^{(i)}$ ≡ ratings of the interaction (e.g. slider [active], facial expression [passive], ...)
  - . $r \in [0,10]$; $r = r_{av} + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$
  - . $r = f(a_j, h)$, where h is the history of previous interactions

User is assumed to provide a rating at each time step

Goal: to learn which action to take and when.

USER MODELING

We introduce three types of user profiles inspired by variations along the "openness" dimension of the five-factor model in psychology [1]

Model 1: The "**conservative**"

Sticks to one option, but appreciates occasional surprises at some frequency. Important parameters are:

- a^*: preferred animation (yielding highest expected reward)
- $\boldsymbol{A_{surp}}$: set of actions suitable for surprises
- $T \sim U([T_{min}, T_{max}])$, optimal sequence length before a surprise

Model 2: The "**consistent but fatigable**"

Enjoys an uninterrupted routine but this routine has to be changed after some time. Important parameters are:

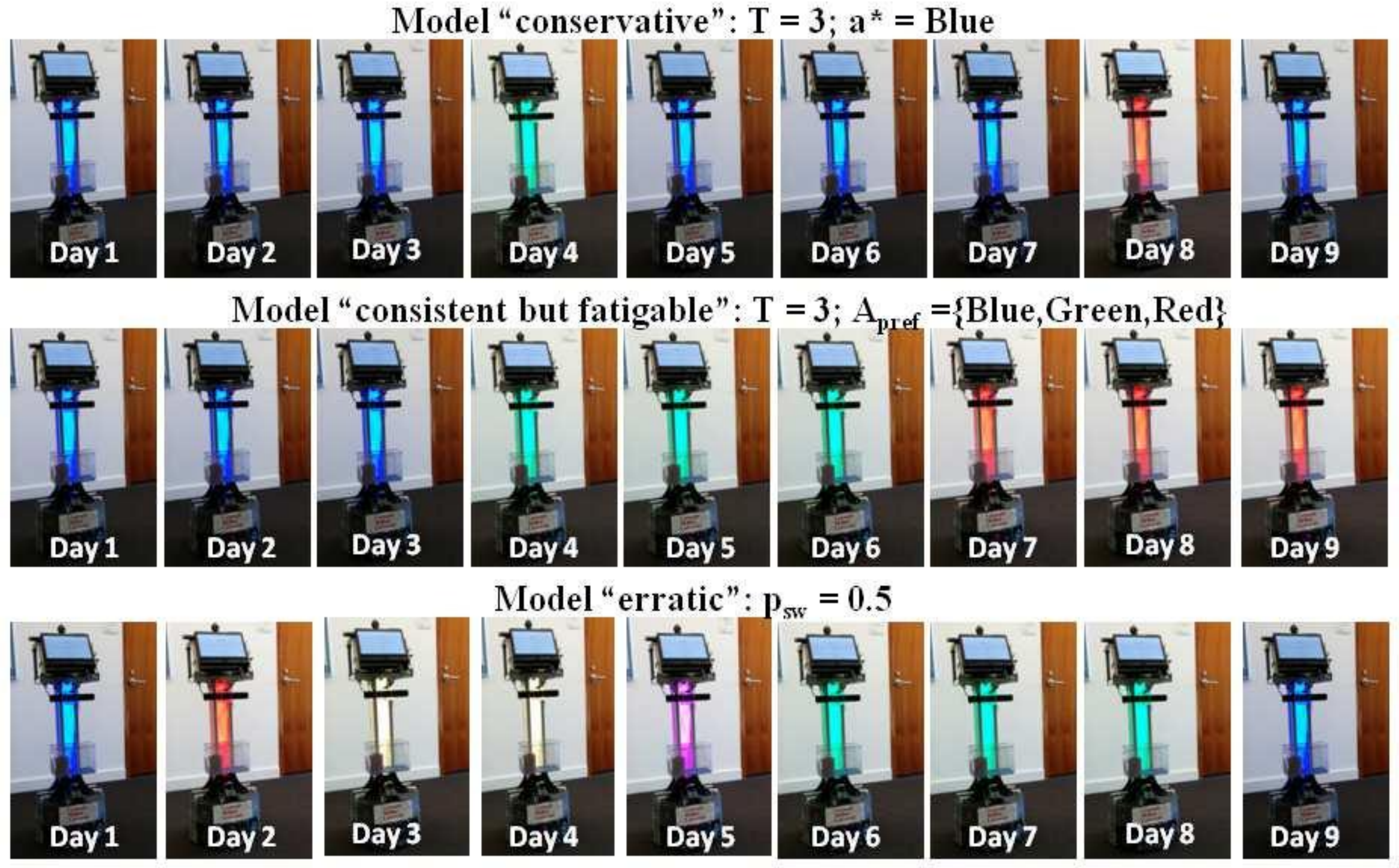
$\boldsymbol{A_{pref}}$: set of preferred actions (yielding expected reward above some threshold)

$T \sim U([T_{min}, T_{max}])$, optimal homogeneous sequence length

Model 3: The "**erratic**"

Mainly interested in change, in both action sets and time-related parameters. The important parameter is:

p_{sw}: fixed probability of desiring a switch at a given time step


Model "conservative": T = 3; a* = Blue

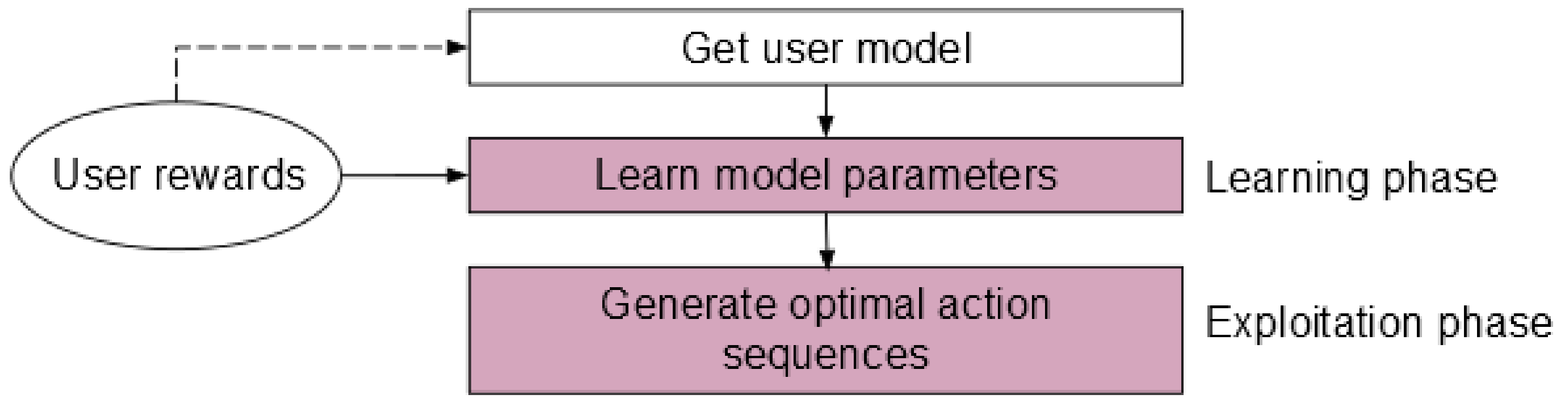
Model "consistent but fatigable": T = 3; A_{pref} ={Blue,Green,Red}

Model "erratic": p_{sw} = 0.5

Fig. 2: Sample preferred of animation sequences for the user models presented

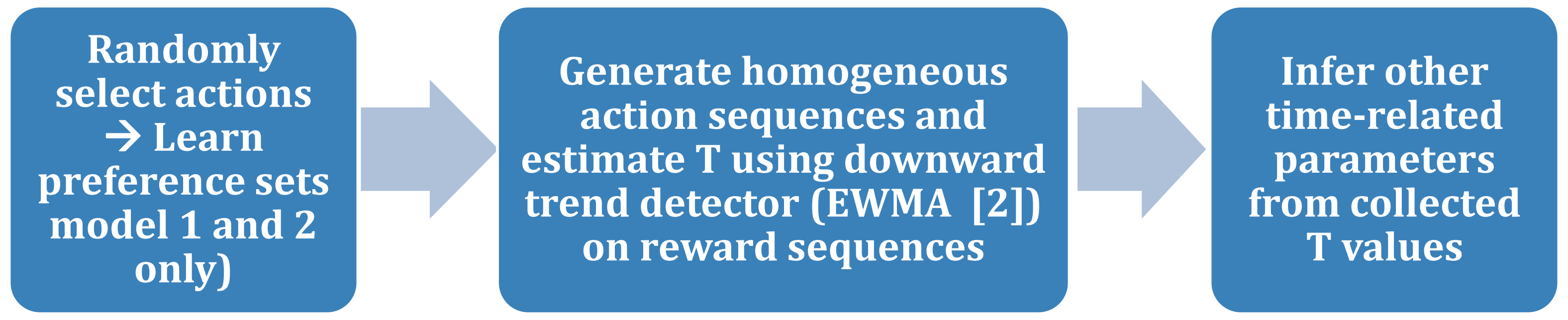
The three models are made flexible enough to cover most of the spectrum of possible preference dynamics.

ONLINE ADAPTATION TO PREFERENCE DYNAMICS



In this work, we assume the user model type is given and we are interested in learning the model parameters in order to generate optimal action sequences.

Learning Procedure


Randomly select actions → Learn preference sets model 1 and 2 only)

Generate homogeneous action sequences and estimate T using downward trend detector (EWMA [2]) on reward sequences

Infer other time-related parameters from collected T values

Sequence Generation (Exploitation Phase)

Once the model parameters are known, the agent can generate optimal sequences as follows:

•Model 1: Draw T uniformly in $[T_{min}, T_{max}]$ to interrupt a^* sequences after *T* steps with a surprise action in $\boldsymbol{A_{surp}}$.

•Model 2: Same but change to another action in $\boldsymbol{A_{pref}}$ after T steps.

•Model 3: Switch to another action with probability p_{sw} at each step.

SIMULATION RESULTS

Our algorithm was tested on three simulated users belonging respectively to each of the three models. After a few time steps, the rewards are maintained to high values robustly.

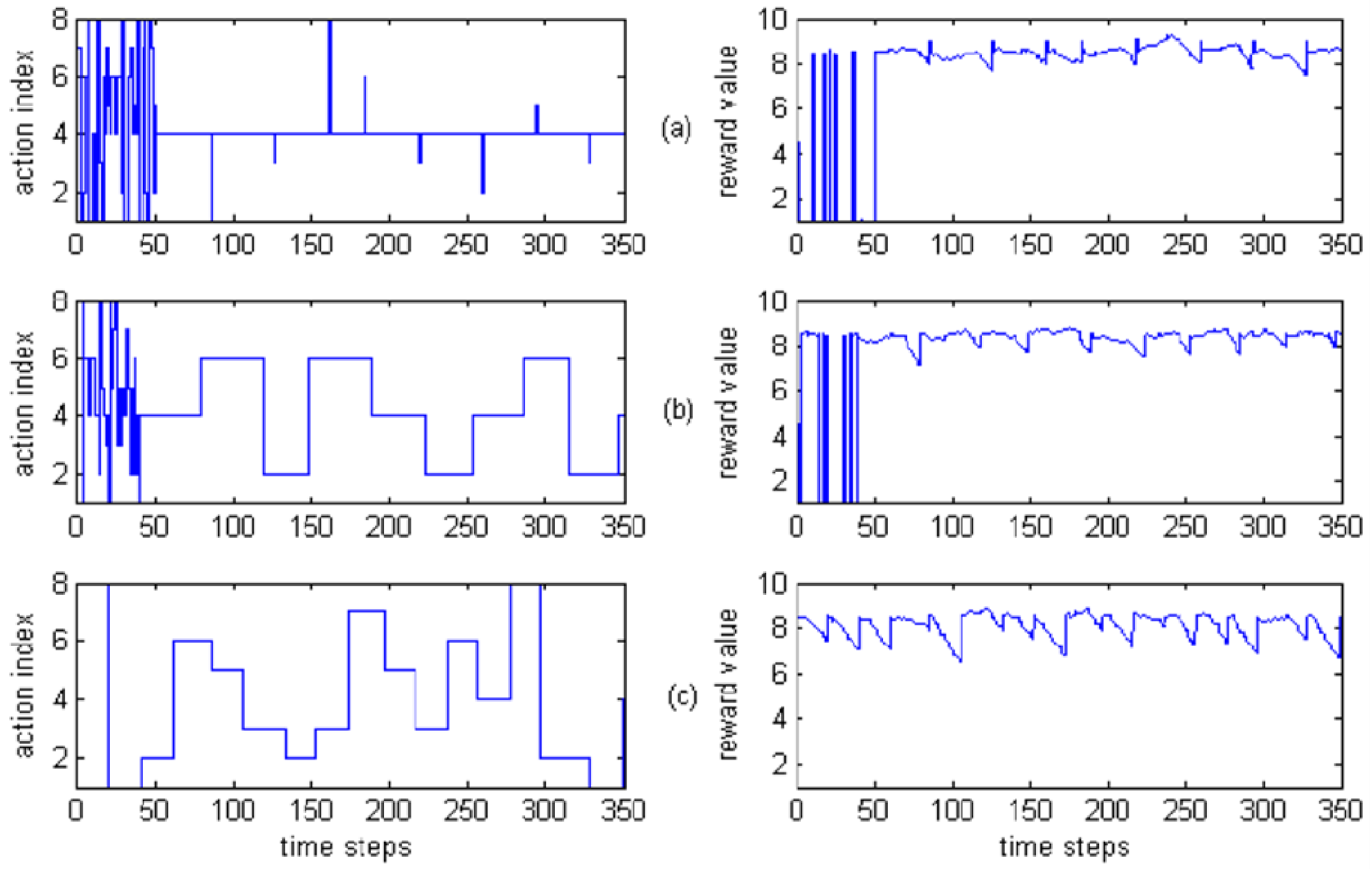


Fig. 3: Simulation results showing action sequences and corresponding reward sequences from a simulated user belonging to models: (a) "conservative", (b) "consistent but fatigable" and (c) "erratic"

FUTURE WORK

- Use a distance metric for large action sets in order to accelerate the learning (e.g. for lights it might be in the color/speed space)
- Evaluate our algorithm's performance with real users

Acknowledgements

This research was partially supported by the National Science Foundation (NSF) and the Fundação para a Ciência e a Tecnologia (FCT)

References

[1] Goldberg, Lewis R. "An alternative" description of personality": the big-five factor structure." Journal of personality and social psychology 59.6 (1990): 1216.

[2] J. Lucas, M. James, M. Saccucci. "Exponentially weighted moving average control schemes: properties and enhancements." Technometrics 32.1, 1990.