FaiR-IoT: Fairness-aware Human-in-the-Loop Reinforcement Learning for Harnessing Human Variability in Personalized IoT

Salma Elmalaki
University of California, Irvine

UCIrvine

Salma Elmalaki
Assistant Professor of Teaching
Electrical Engineering and Computer Science
UC Irvine

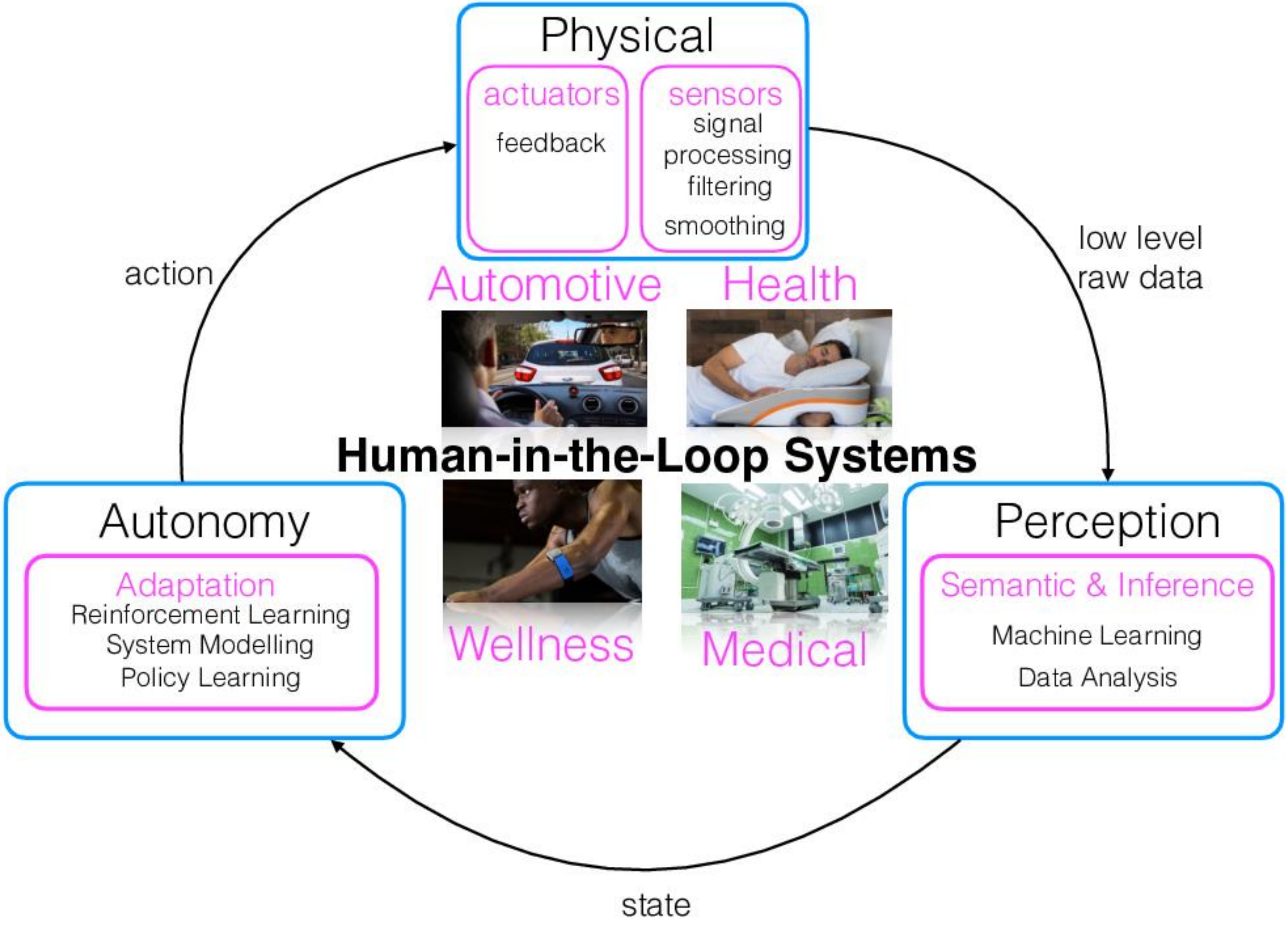
https://faculty.sites.uci.edu/elmalaki/

References

- Elmalaki, S. (2021a, May 20). *FaiR-IoT: Fairness-aware Human-in-the-Loop Reinforcement Learning for Harnessing Human Variability in Personalized IoT.* Internet of Things Design and Implementation (IoTDI) 2021. https://youtu.be/T7G7cbEDZ1c?list=PLTXOFWg43hoVPVO4z9VzLE38Q3oYrcplS&t=5523
- Elmalaki, S. (2021b). FaiR-IoT: Fairness-aware Human-in-the-Loop Reinforcement Learning for Harnessing Human Variability in Personalized IoT. *Proceedings of the International Conference on Internet-of-Things Design and Implementation*, 119–132. https://doi.org/10.1145/3450268.3453525
- Elmalaki, S., Tsai, H.-R., & Srivastava, M. (2018). Sentio: Driver-in-the-Loop Forward Collision Warning Using Multisample Reinforcement Learning. *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 28–40. https://doi.org/10.1145/3274783.3274843

Agenda

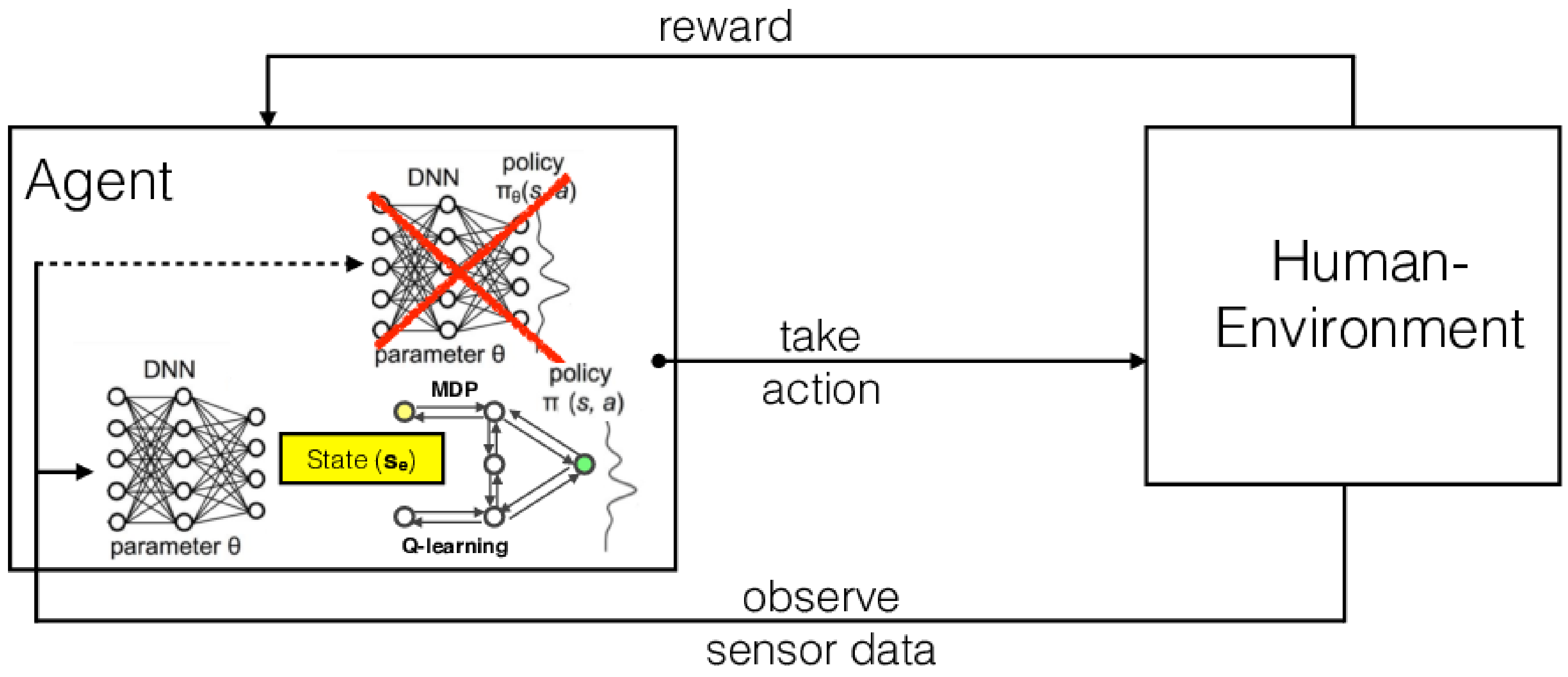
1. Overview
   - **Variability and Fairness**
   - Experiments and Results
2. Positive Points
3. Negative Points
4. Questions/Discussion


Physical
actuators
feedback
sensors
signal
processing
filtering
smoothing
low level
raw data
action
Automotive
Health
Human-in-the-Loop Systems
Autonomy
Adaptation
Reinforcement Learning
System Modelling
Policy Learning
Wellness
Medical
Perception
Semantic & Inference
Machine Learning
Data Analysis
state

Human Variability

1. Intra-human: user wants different things at different times
    - Multisample RL
2. Inter-human: users want different things
    - Governer RL
3. Multi-human: users want conflicting things
    - Mediator RL

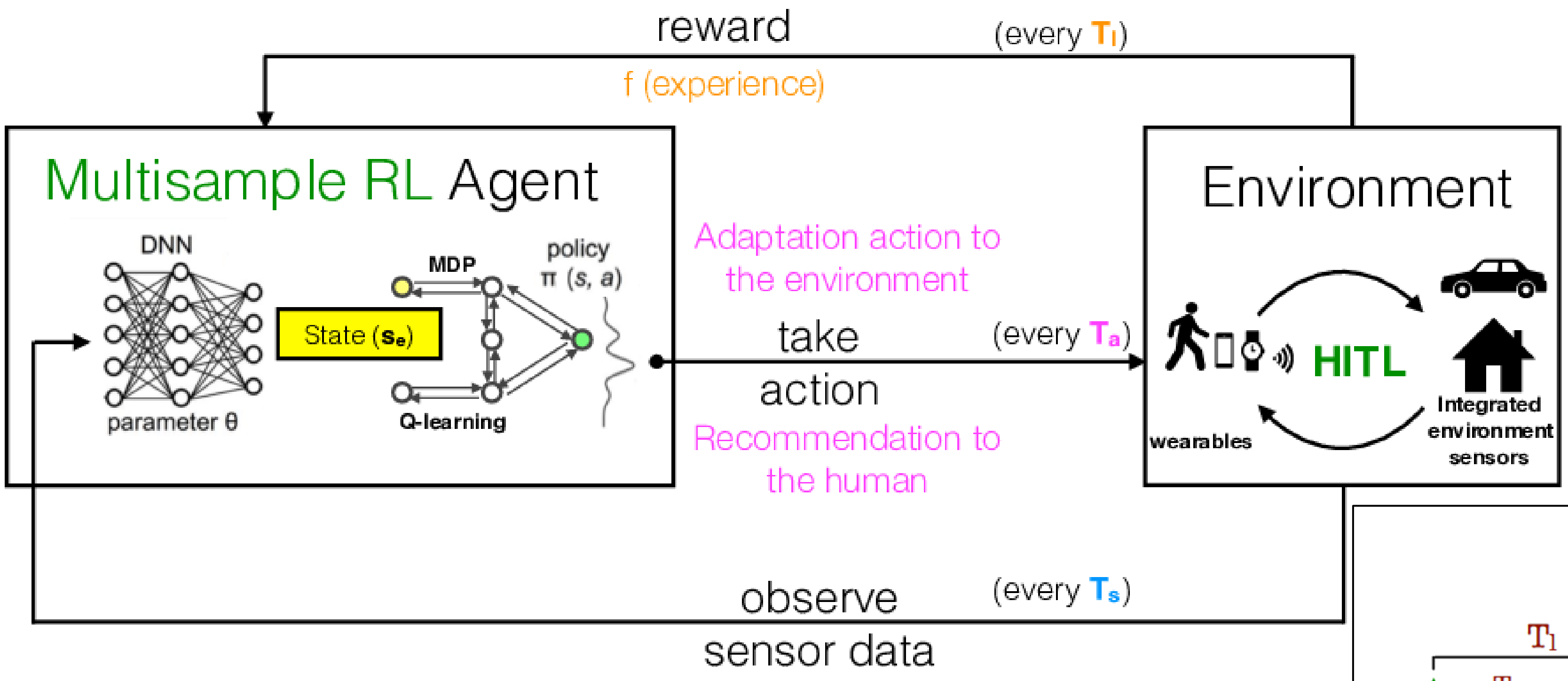
Human-Environment Interaction with Reinforcement Learning



- Traditional RL algorithms assumes:
  - time-invariant rewards.
  - actions impact the environment immediately
  - rewards accrue from corresponding action

Human-Environment Interaction with Multisample Reinforcement Learning

(Sensys'18)



- Environment is observed every T_s samples (Observing Rate)
- Action is taken by the agent every T_a samples (Actuation Rate)
- Reward is calculated after an action by T_l samples (Learning Rate)

$$R_a(s_l, t_l) = \sum_{i=t_l}^{t_l+n} I_a(i) \times R'(s_i)$$

Reward is calculated over a time horizon after action is released

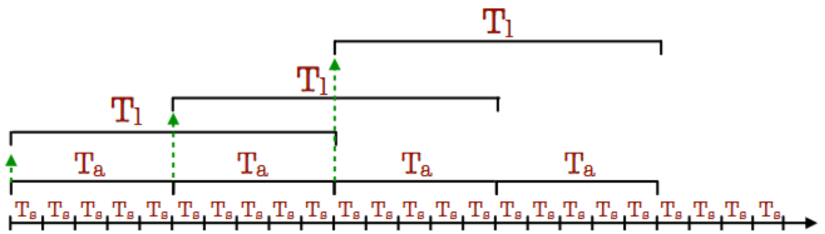
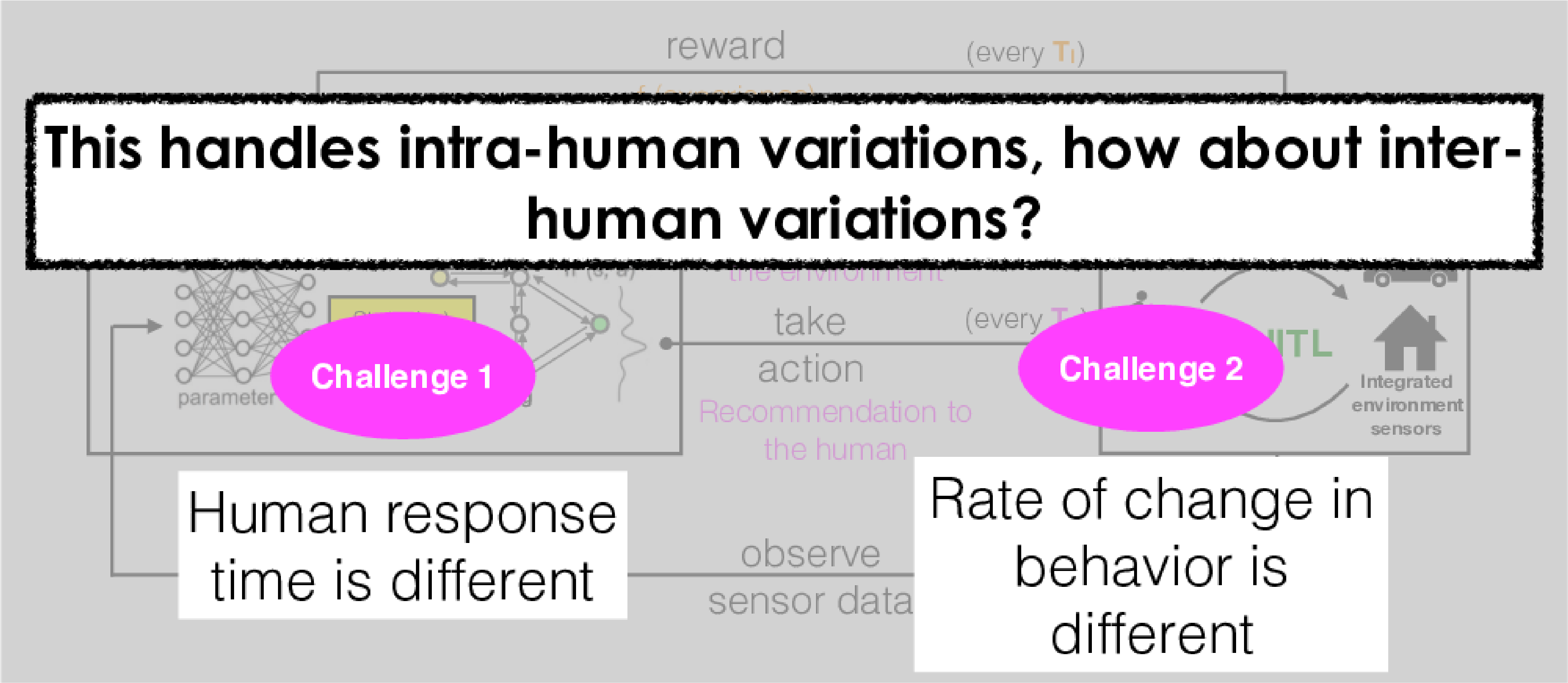


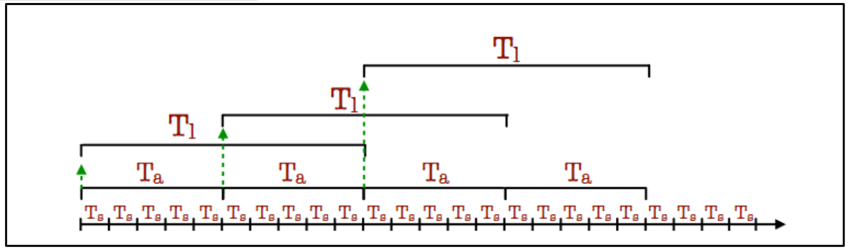
Figure 4: Multisample timescales. An action is taken by the agent every T_a samples. The reward is calculated after an action is taken by a time equals T_l. In this example, $T_a = 5T_s$ and $T_l = 10T_s$, where T_s is the state observing (sensor) rate.

(Elmalaki et al., 2018)

Inter-Human Variation Using Governor RL

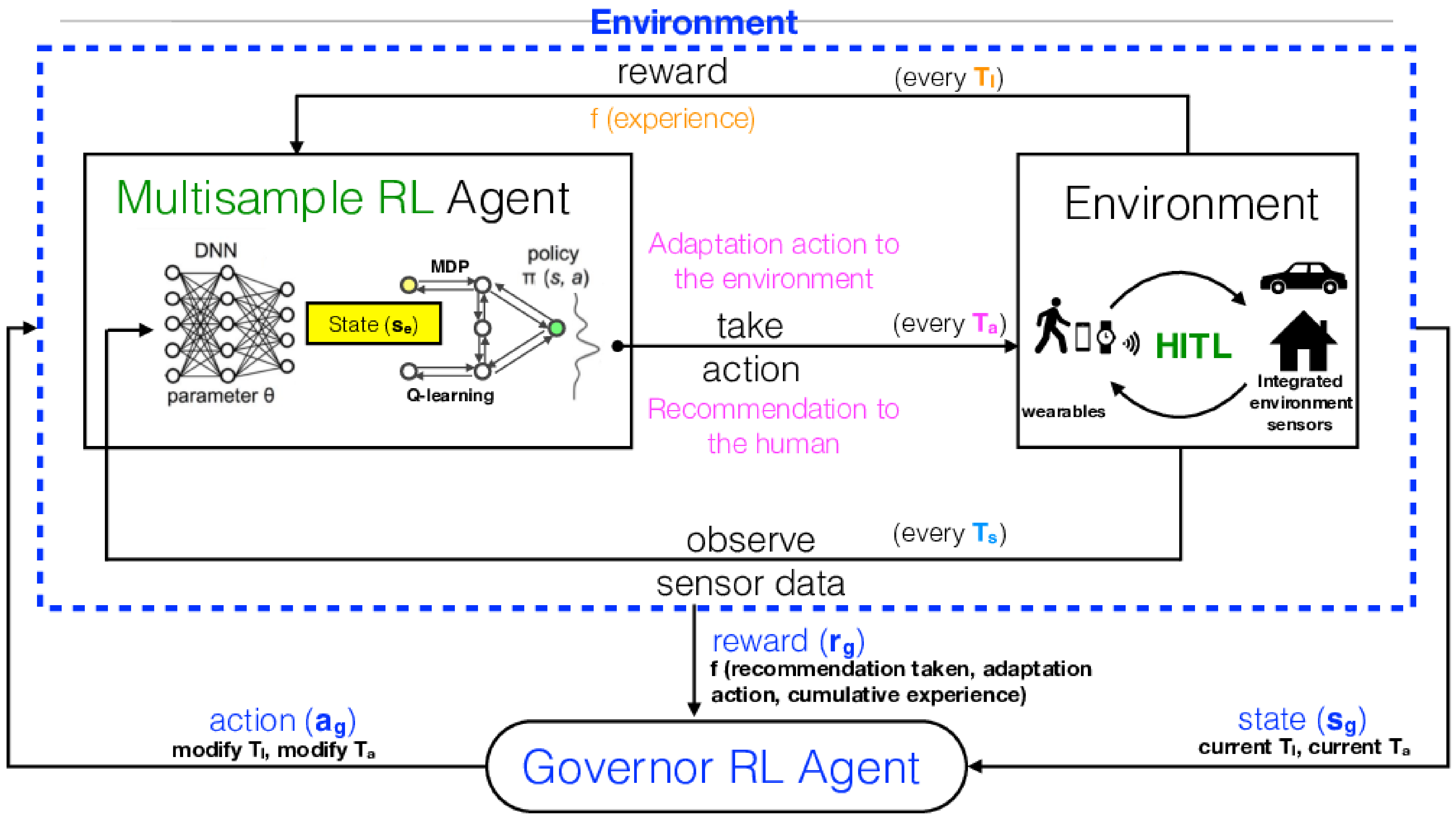


- T_a and T_l can not be fixed and they are different from one human to another
- Action is taken by the agent every T_a samples (Actuation Rate)
- Reward is calculated after an action by T_l samples (Learning Rate)

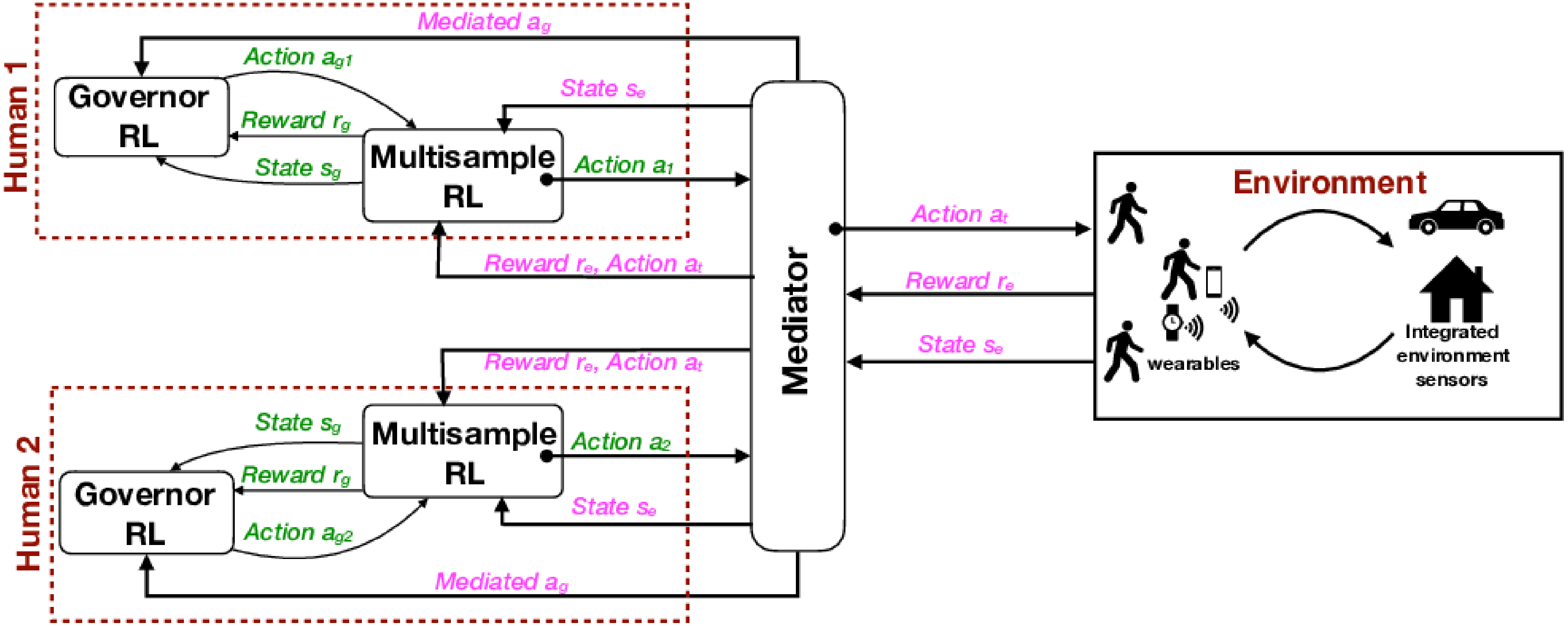


(Elmalaki et al., 2018)

Inter-Human Variation Using Governor RL



Multi-Human Environment

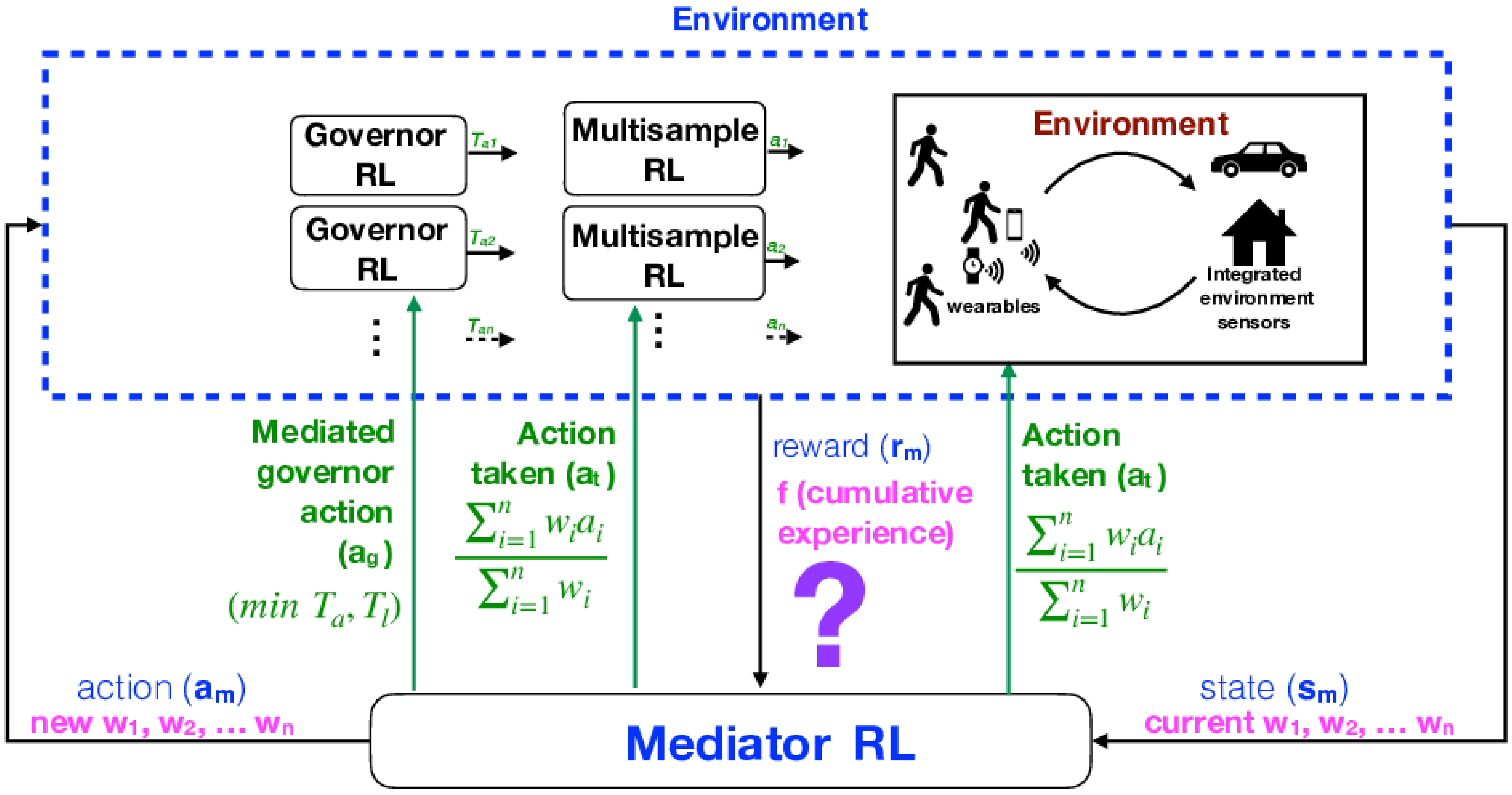


How to design Mediator?

Mediator chooses action:

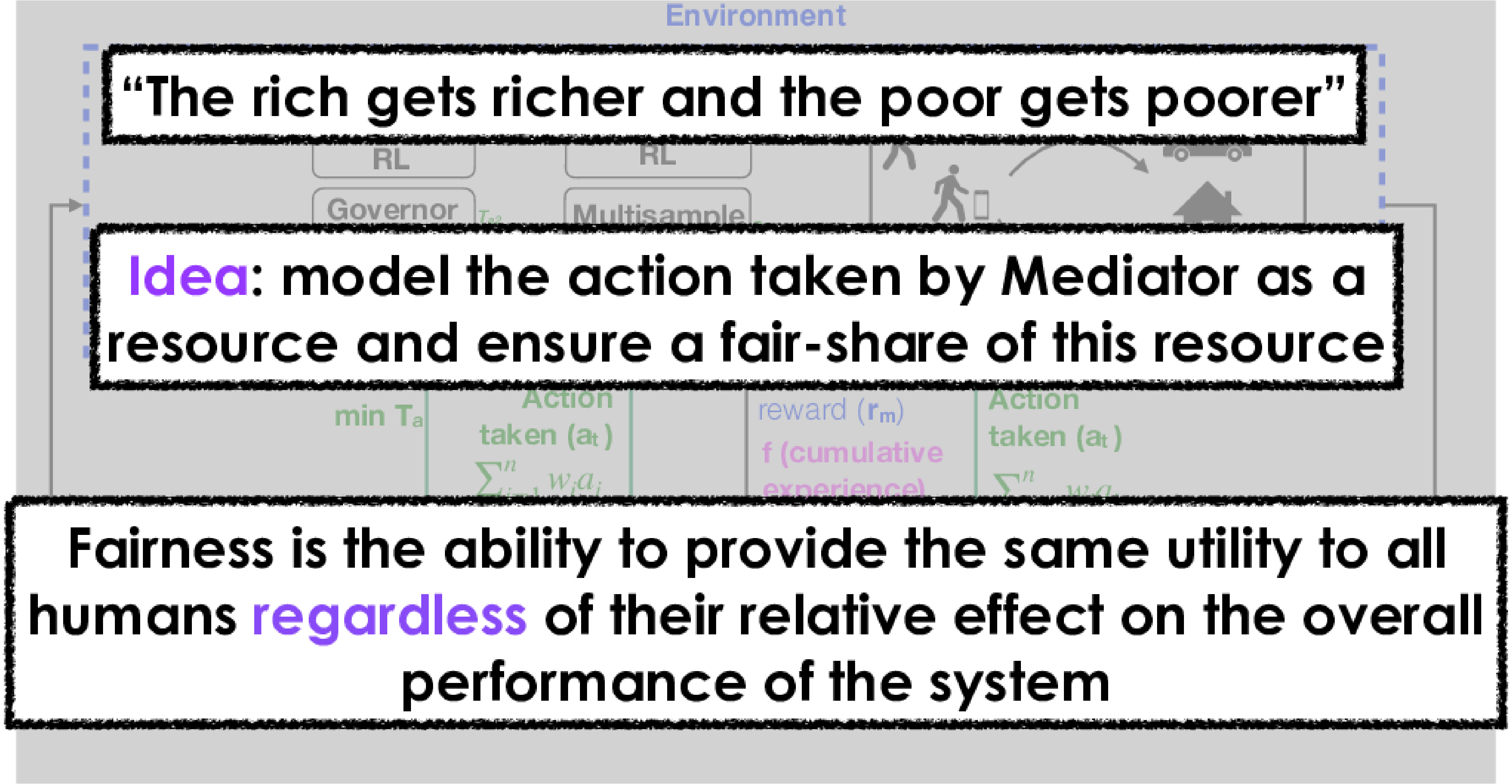
- Minimum Ta, Tl
- Weighted average of suggested actions
  - Mediator learns weights

Mediator RL Design



How to ensure fairness?

Matthew effect of accumulated advantage



“The rich gets richer and the poor gets poorer”

Idea: model the action taken by Mediator as a resource and ensure a fair-share of this resource

Fairness is the ability to provide the same utility to all humans **regardless** of their relative effect on the overall performance of the system

Fairness as a notion of fair share of utility

Fairness is the ability to provide the same utility to all humans regardless of their relative effect on the overall performance of the system

$$u_{h_t} = \frac{1}{t}\sum_{j=0}^{t}\frac{j}{t}w_{h_j}$$

u_{h_t} is the average weight assigned by the Mediator RL for a particular human h over a time horizon *[0:t]*

j/t emphasizes recent weights

Measure the fairness using coefficient of variations of the human utilities

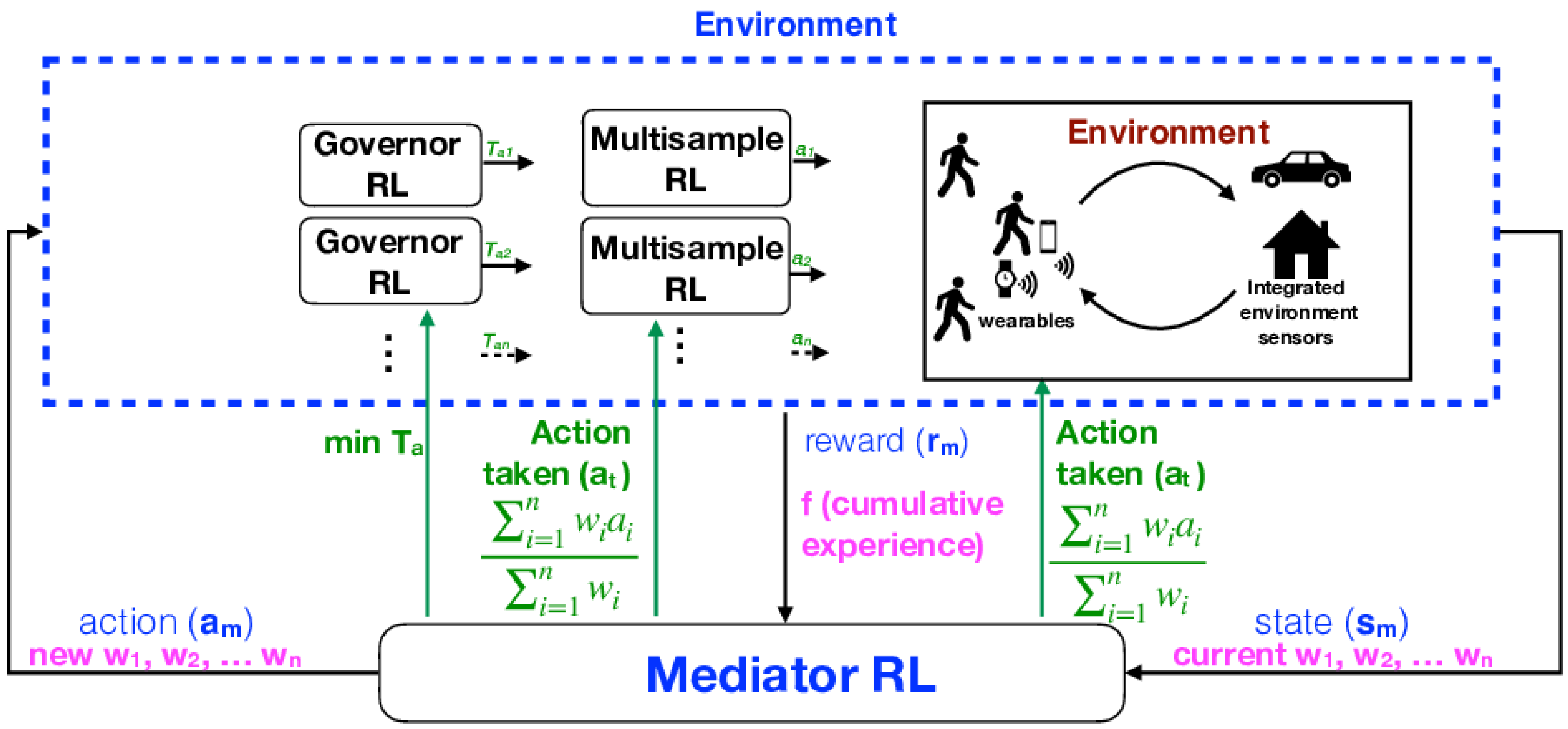
$$cv = \sqrt{\frac{1}{n-1}\sum_{h=1}^{n}\frac{(u_h - \bar{u})^2}{\bar{u}^2}}$$

$\bar{u}$ is the average utility of all humans

Coefficient of variation is ratio of standard deviation to mean

The Mediator RL is said to be more fair if and only if the *cv* is smaller.

Fairness-aware Reinforcement Learning

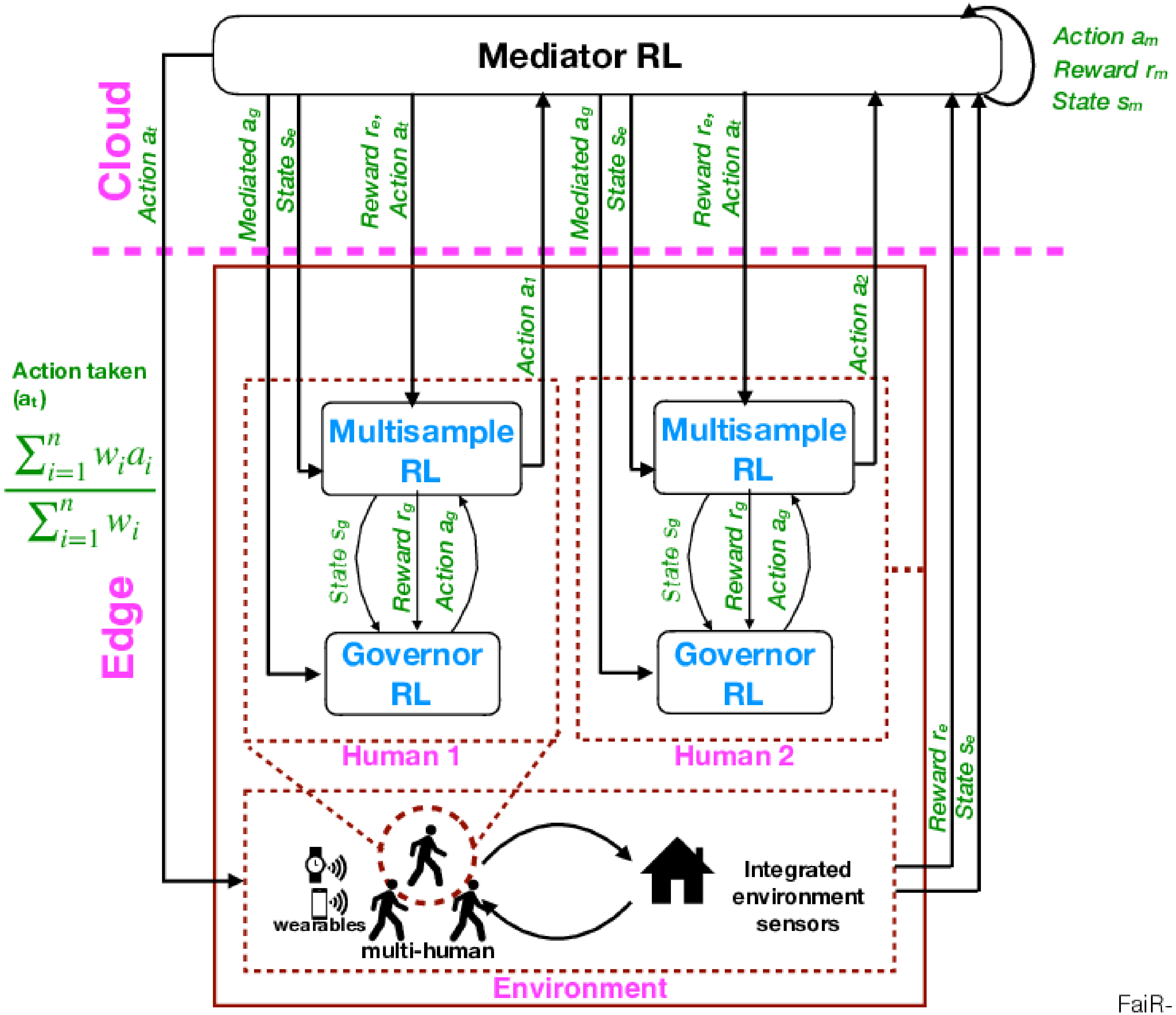


$$r_m \leftarrow R_{\mathcal{M}}(s_m, a_m) = (1-\zeta)\mathcal{W}(p_{s'}, p_s) + \zeta\mathcal{F}(cv_{s'_m}, cv_{s_m})$$

ζ: Fairness scale [0:1]

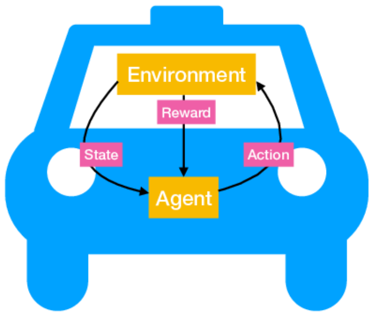
- ps: performance at state s
- cv_s: coefficient of variation at state s
  - I don't know what subscript m means, it didn't appear in paper (Elmalaki, 2021, 124).
- W: performance function
- F: difference between coefficients of variation

FaiR-IoT Architecture



Agenda

1. Overview
   - Variability and Fairness
   - **Experiments and Results**
2. Positive Points
3. Negative Points
4. Questions/Discussion



Advanced Driver Assistance System (ADAS)

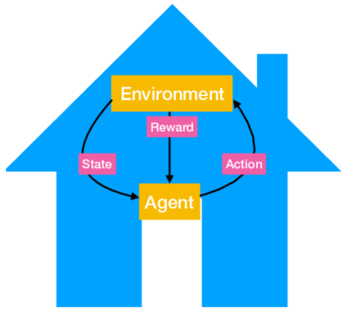
Drivers H1, H2, and H3 with moderate, aggressive, and slow behavior

1: Inter-Human Variability: Different drivers

- Best Tl, Ta for H2 and H3, second best for H1

2: Intra-Human Variability: Driver changes behavior

- Adapted to changes (H1, H2) and (H1, H3)
- Why not switch between H2 and H3?



Thermal System

Residents H1, H2, H3 with decreasing levels of activity

3: Inter-Human Variability: Single person in house

- Best Tl, Ta for H_1, second best Tl, Ta for H2

4: Multi-Human Variability: Multiple people in house

- Only considered H3 (weighted average with weights (0, 0, 1))

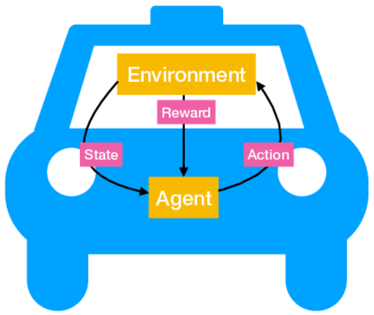
5: Multi-Human Variability + Fairness: Multiple people in house

- Trades performance for fairness compared with 4

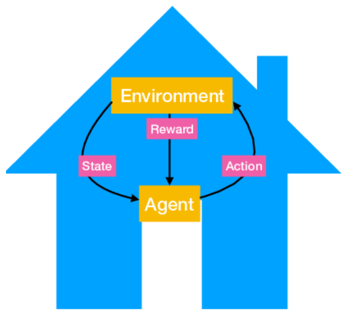
6: Comparison with Fixed Point

- Performance improved 41.7% and 58.96% compared to 70F and 76F set point

- Both simulations due to pandemic
- ADAS with Multisample RL implemented and tested with human subjects in Sentio paper (Elmalaki et al., 2018).



Advanced Driver Assistance System (ADAS)



Thermal System

Residents H1, H2, H3 with decreasing levels of activity

3: Inter-Human Variability: Single person in house

- Best Tl, Ta for H_1, second best Tl, Ta for H2

4: Multi-Human Variability: Multiple people in house

- Only considered H3 (weighted average with weights (0, 0, 1))

5: Multi-Human Variability + Fairness: Multiple people in house

- Trades performance for fairness compared with 4

6: Comparison with Fixed Point

- Performance improved 41.7% and 58.96% compared to 70F and 76F set point

Drivers H1, H2, and H3 with moderate, aggressive, and slow behavior

1: Inter-Human Variability: Different drivers

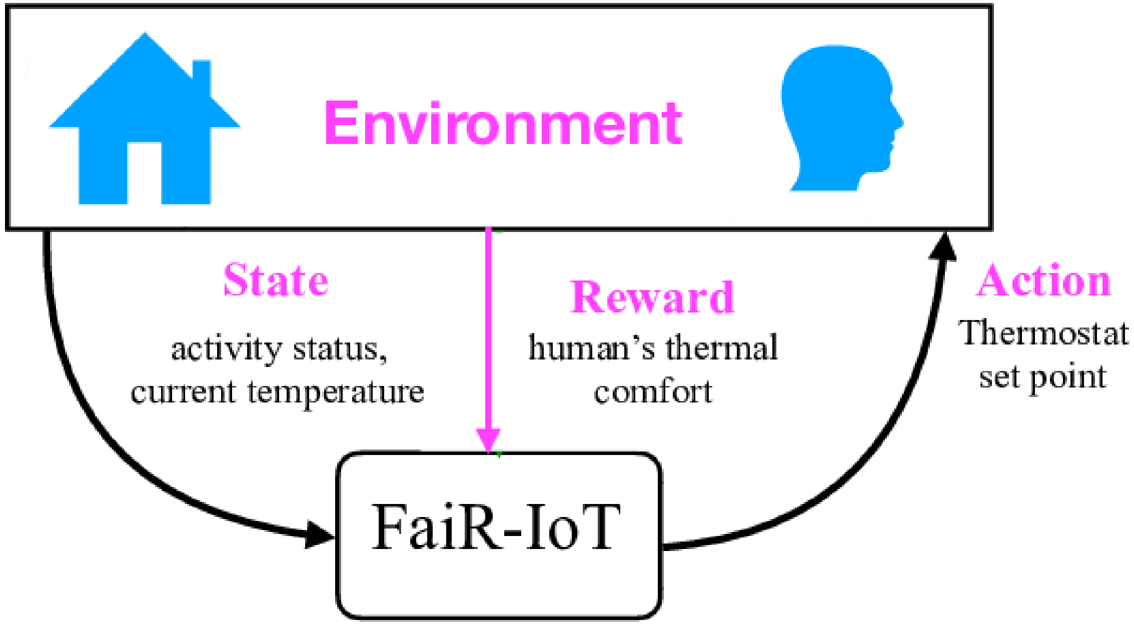
- Best Tl, Ta for H2 and H3, second best for H1

2: Intra-Human Variability: Driver changes behavior

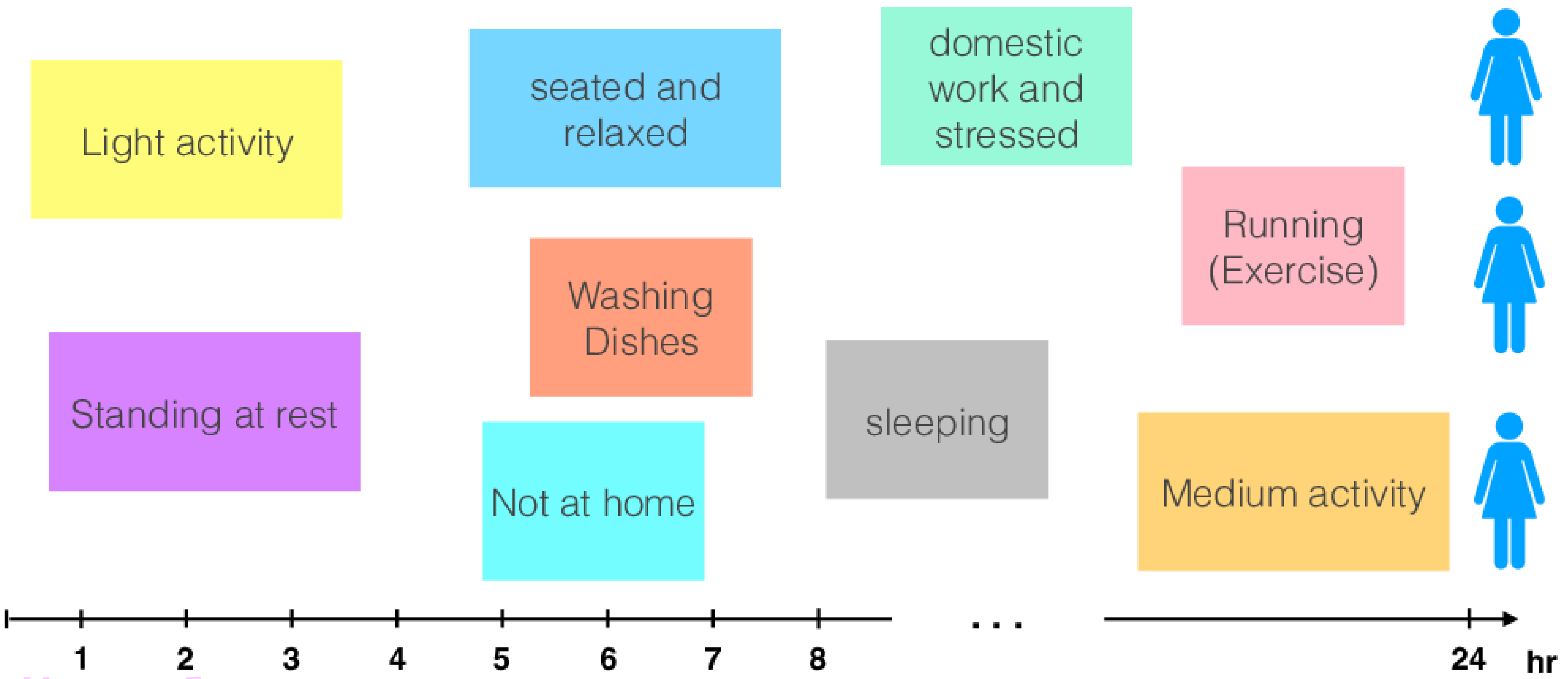
- Adapted to changes (H1, H2) and (H1, H3)
- Why not switch between H2 and H3?

- Both simulations due to pandemic
- ADAS with Multisample RL implemented and tested with human subjects in Sentio paper (Elmalaki et al., 2018).

Example: Human-in-the-Loop Smart House: Thermal System



Personalization of Thermal Home System using FaiR-IoT

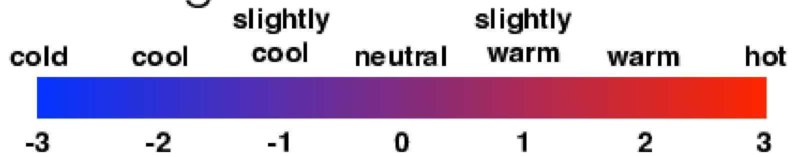


Human Data:

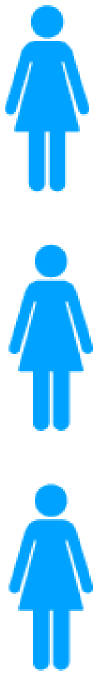
Simulation of three human occupants across a period of 5 days for a pattern of 9 activities

Performance metrics:

Change of PMV across time



Human-in-the-Loop of Thermal Home System



Human Modeling

Heat source with heat flow that depends on:

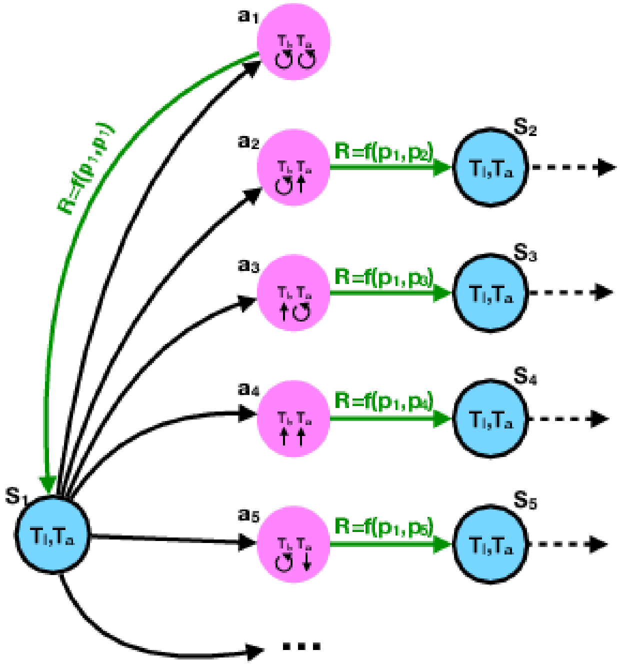
- the average exhale breath temperature (EBT)
- the respiratory minute volume (RMV)

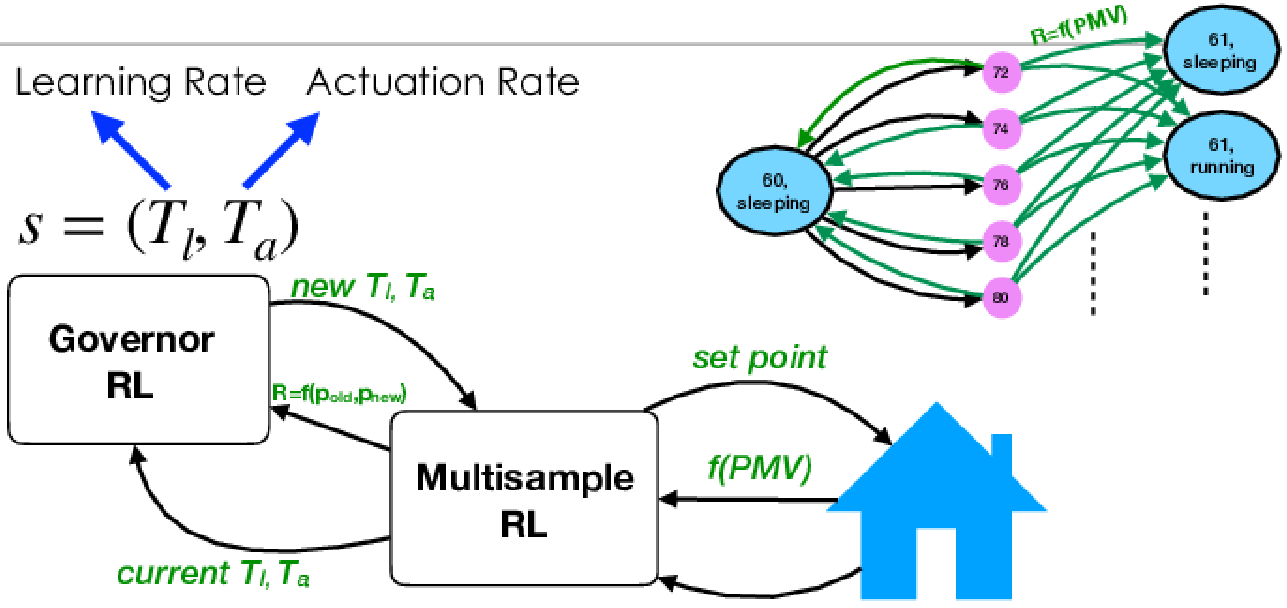
These two parameters highly dependent on human activity

House Modeling

- Thermodynamic model of a house
- Design FaiR-IoT to control the thermostat of the house

Intra- Inter-Human Variability with Multisample RL and Governor RL





Different values of Tl and Ta may give better thermal comfort for different humans (inter-human variability)

$T_s = 6\ min$

$T_l \in [10T_s,\ 15T_s,\ 20T_s,\ 30T_s]$

$T_a \in [2T_s,\ 4T_s,\ 6T_s,\ 8T_s,\ 10T_s,\ 12T_s,\ 14T_s,\ 16T_s]$

$T_l \geq T_a$

$$\overline{PMV}_{MA_s} = \frac{1}{n}\sum_{i=0}^{n-1}\left|pmv_i\right|$$

$$\bar{\sigma}_{MA_s} = \frac{1}{n}\sum_{i=0}^{n-1}\sigma_i$$

$$p_s = \theta_1\overline{PMV}_{MA_s} + \theta_2\bar{\sigma}_{MA_s}$$

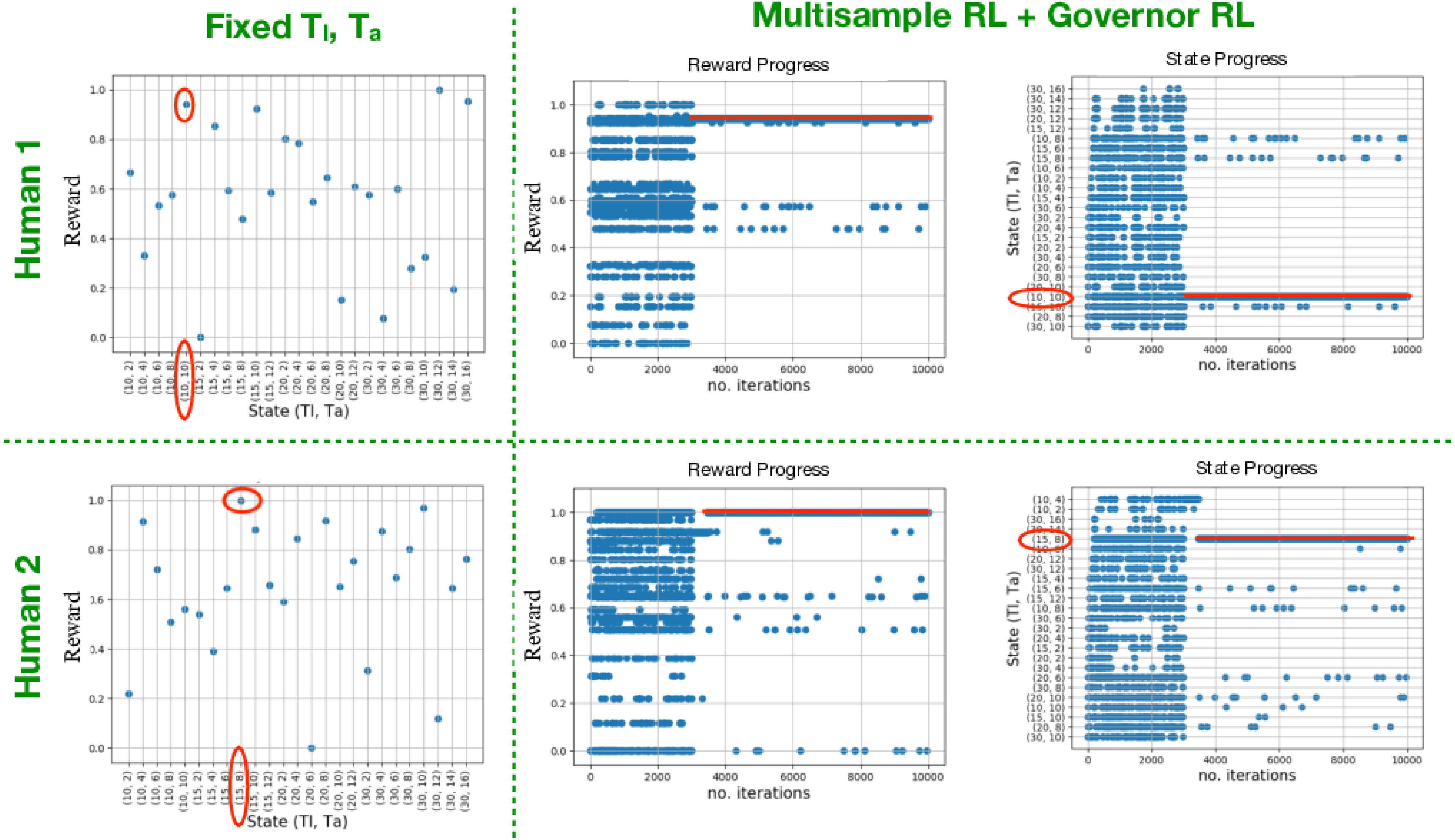
$s = (a, t)$

activity

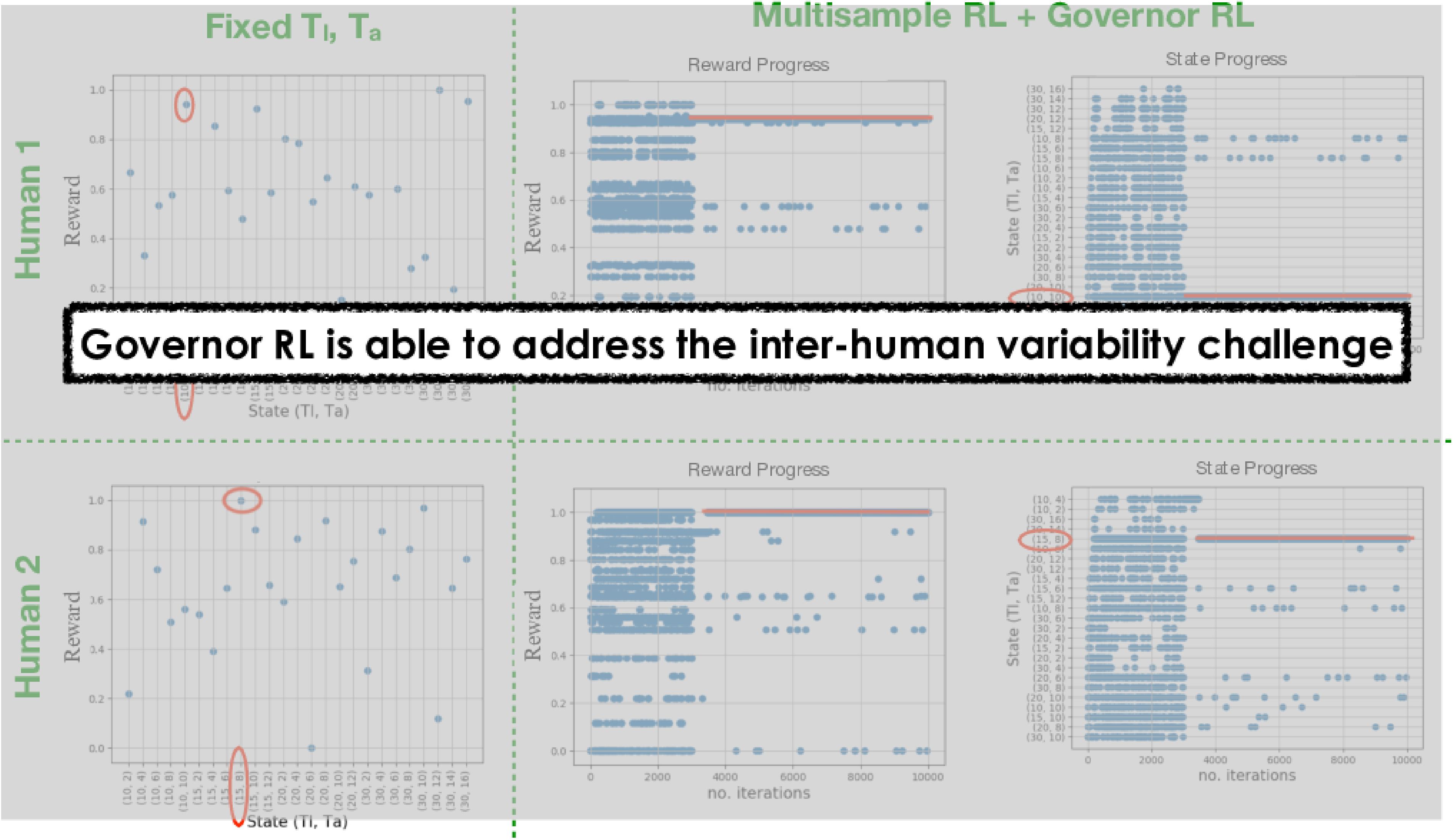
current indoor temperature $[60F,\ 80F]$

Each human may response differently to HVAC set point based on the current activity

Intra- Inter-Human Variabilities

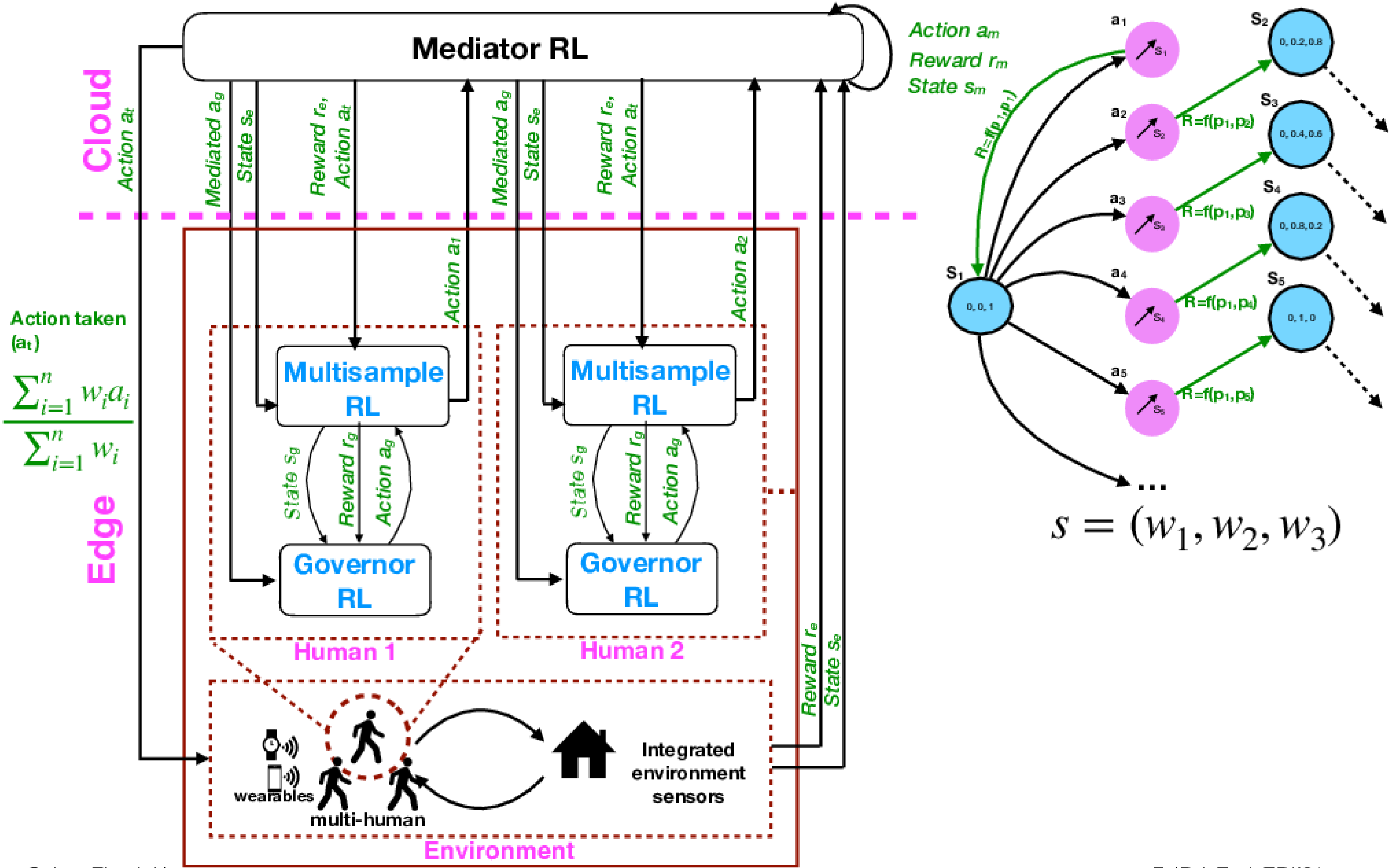


Intra- Inter-Human Variabilities



Governor RL is able to address the inter-human variability challenge

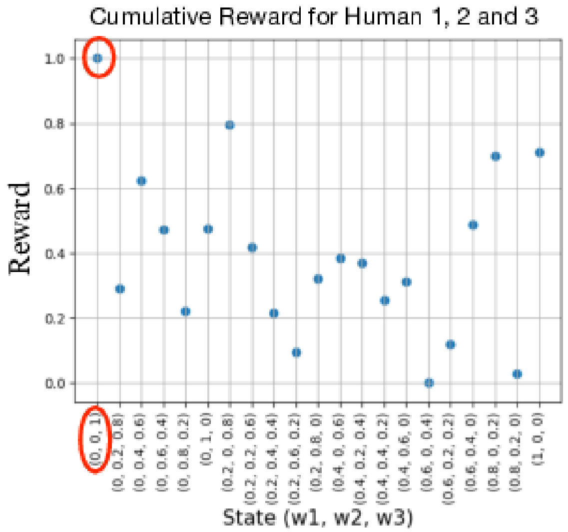
Mediator RL for Multi-Human Variabilities



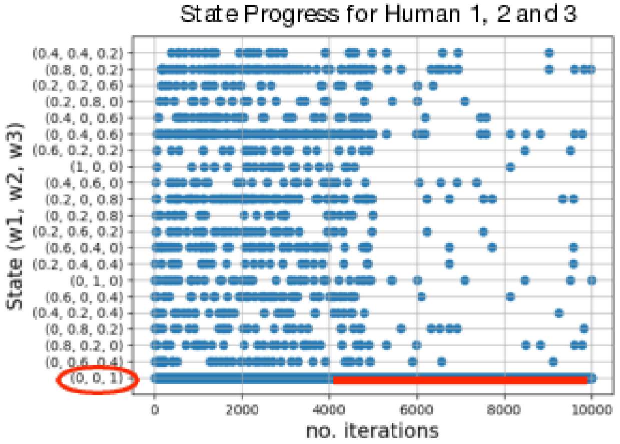
4: Multi-Human Variability: Multiple people in house

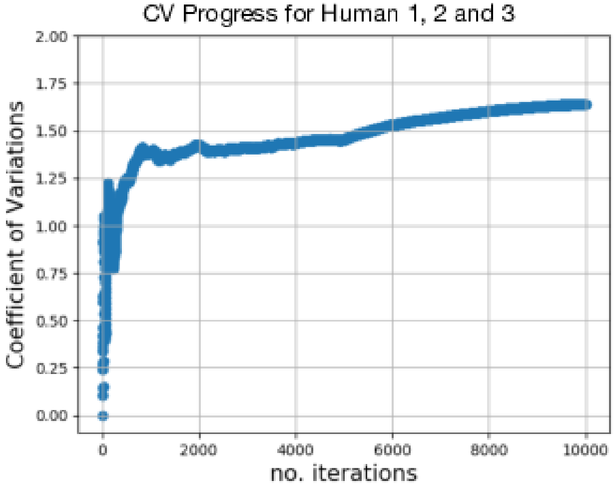
Multi-Human Variabilities - No Fairness

Fixed w_1, w_2, w_3



Mediator RL without Fairness Regulator

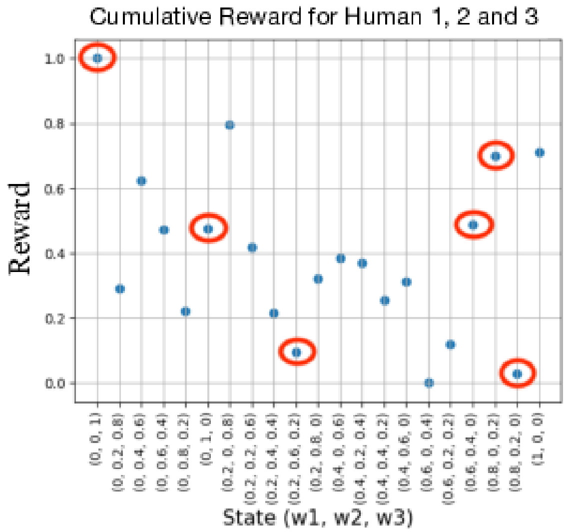




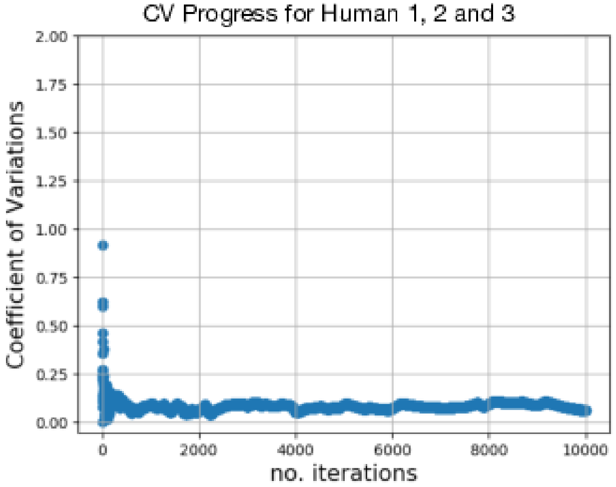
5: Multi-Human Variability + Fairness: Multiple people in house

Multi-Human Variabilities - With Fairness

Fixed w_1, w_2, w_3



Mediator RL with Fairness Regulator



The lower value of cv indicates higher fairness. In particular, the cv has been improved by 1.5 orders of magnitude through integrating the fairness measure $\mathscr{F}$ in the reward function of the Mediator RL

$$r_m \leftarrow R_{\mathscr{M}}(s_m, a_m) = (1-\zeta)\mathscr{W}(p_{s'}, p_s) + \zeta\mathscr{F}(cv_{s'_m}, cv_{s_m})$$

Positive Points

- Implemented and tested FaiR-IoT
- Car scenario modeled on human responses
- Elegance of using RL to optimize prameters

Negative Points

- Only tested Thermal system in simulation
  - Assumes only source of heat are humans and heater
- Doesn’t compare against other smarter approaches, only fixed point
- Fairness requires actions that can be averaged
- Assumes we can easily measure human comfort metric (PMV)

Questions

- Would it be possible to extend fairness to cases where you can't average actions?
- In an implementation of the Thermal System, how could we measure performance?
- Would it be possible to implement this approach on off the shelf equipment?