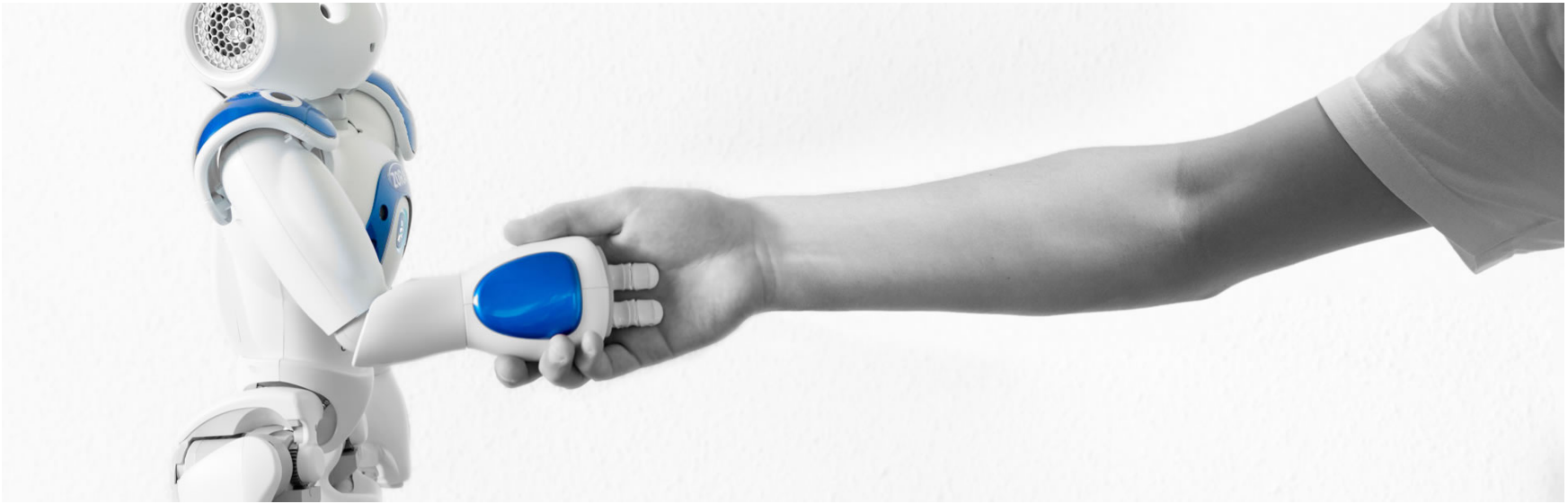
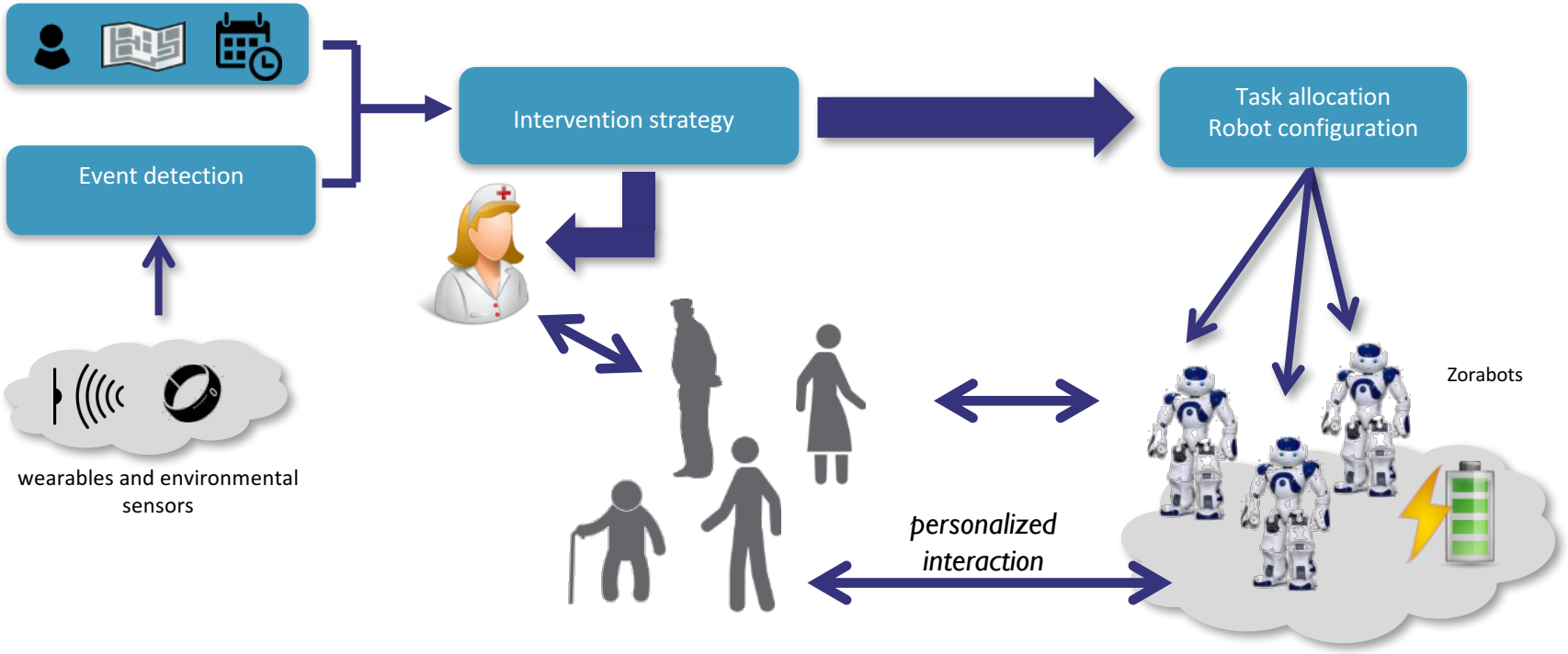
imec

SELF-LEARNING ALGORITHMS FOR THE PERSONALIZED INTERACTION WITH PEOPLE WITH DEMENTIA

BRAM STEENWINCKEL

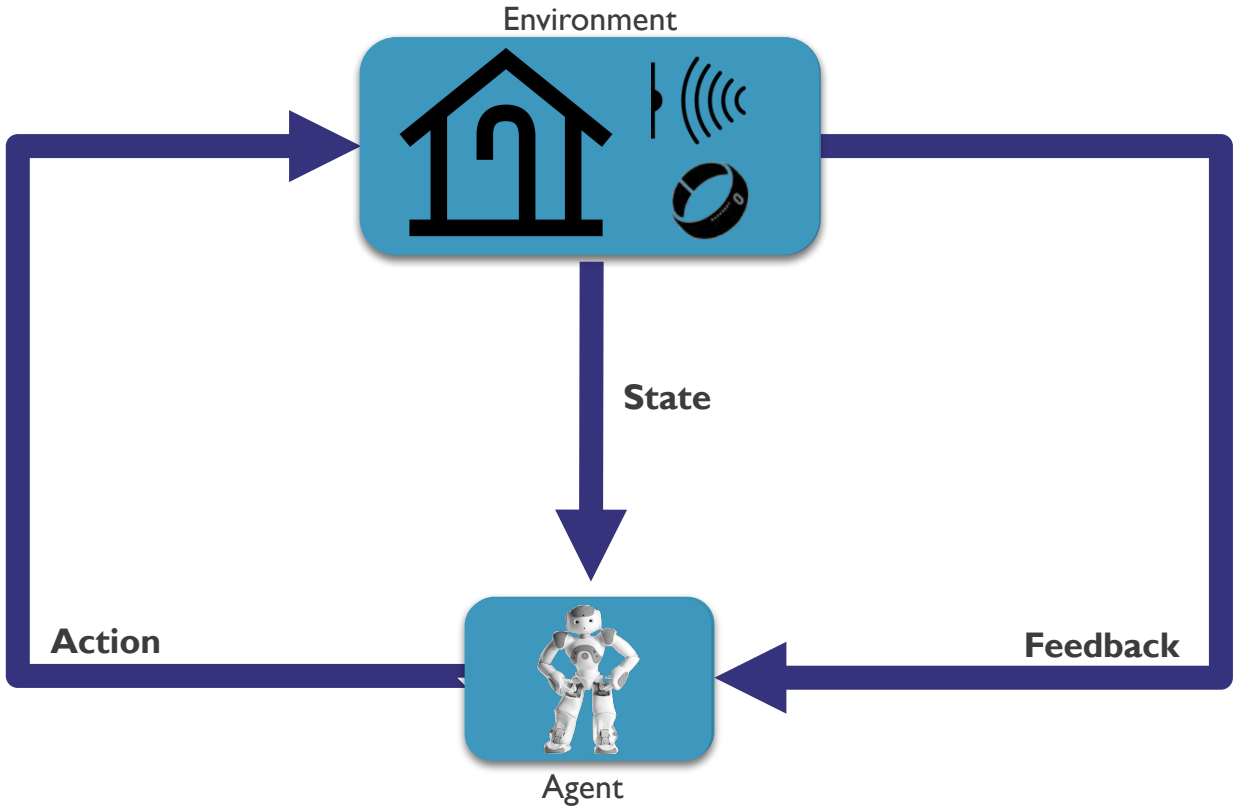


WONDER PROJECT



LEARNING ALGORITHMS

LEARNING



AGENT

– Normal Agent

Datastructure:

estimates

1	2	3	4
0.66	0.4	0.3	0.2

actions

1	2	3	4
5	4	1	7

Update function:

$$g(a) = {}^{1}/_{N(a)}$$

$$estimate_t(a) \mathrel{+}= g(a) * (r_t - estimate_{t-1}(a))$$

Example:

Execution of action 1

Reward value r_t: 0.9

$N(1)$: 5 → g(1) : 0.2

$estimate_{t-1}(1)$: 0.6

$estimate_t(1)$ += 0.2 *(0.9-0.6)

$estimate_t(1)$ = 0.66

AGENT

- Normal Agent
- Gradient Agent

Datastructure:

Preferences			
0.82	0.52	0.52	0.52

Action probabilities $\pi_t(a)$			
0.25	0.25	0.25	0.25

Update function:

$$preference_{t+1}(a) = preference_t(a) + \alpha(r_t - \overline{r_t})\left(1 - \pi_t(a)\right)$$
$$preference_{t+1}(a_n) = preference_t(a_n) + \alpha(r_t - \overline{r_t})\,\pi_t(a_n), \forall a_n \neq a$$

$$\pi_t(a) = \frac{e^{preference_t(a)}}{\sum_{x=1}^{k} e^{preference_t(x)}}$$

Example:

Execution of action 1
Reward value r_t: 0.9

Average reward (baseline) $\overline{r_t}$: 0.6
α: 1.0

$preference_t(1)$: 0.6 and $\pi_t(1)$: 0.25

$preference_{t+1}(1)$ = 0.6 +1.0*(0.9-0.6)*(1-0.25) = 0.825
$preference_{t+1}$(not 1) = 0.6 -1.0*(0.9-0.6)*(0.25) = 0.525

$\pi_t(1)$ = e^(0.825)/(e^(0.825)+3*e^(0.525)) = 0.31
$\pi_t(not\ 1)$ = e^(0.525)/(e^(0.825)+3*e^(0.525)) = 0.22

AGENT

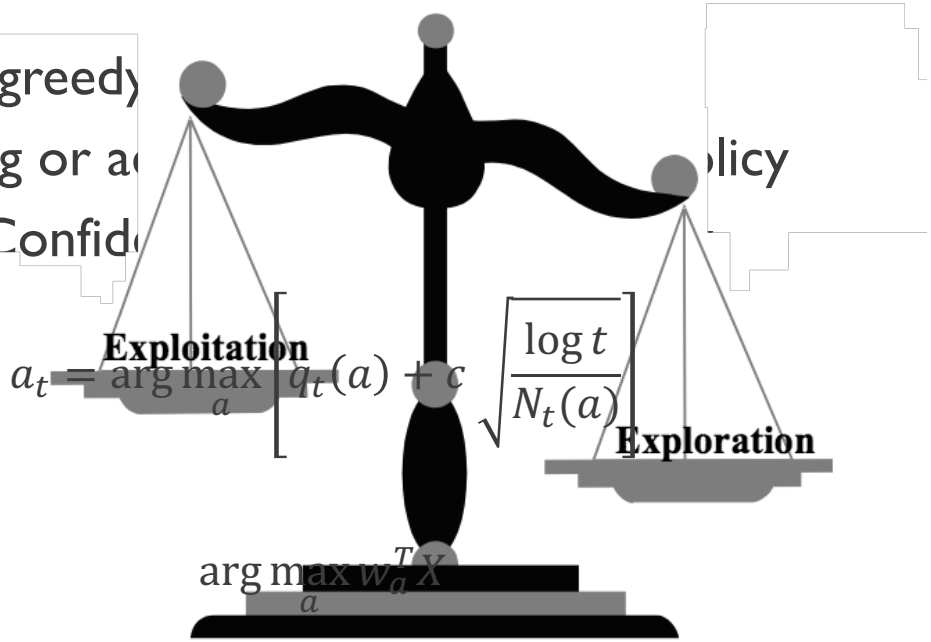
- Normal Agent
- Gradient Agent
- Contextual Agent

$$\mathbb{E}_a(\, r \mid X) = \mathrm{w}_\mathrm{a}^T\, X$$

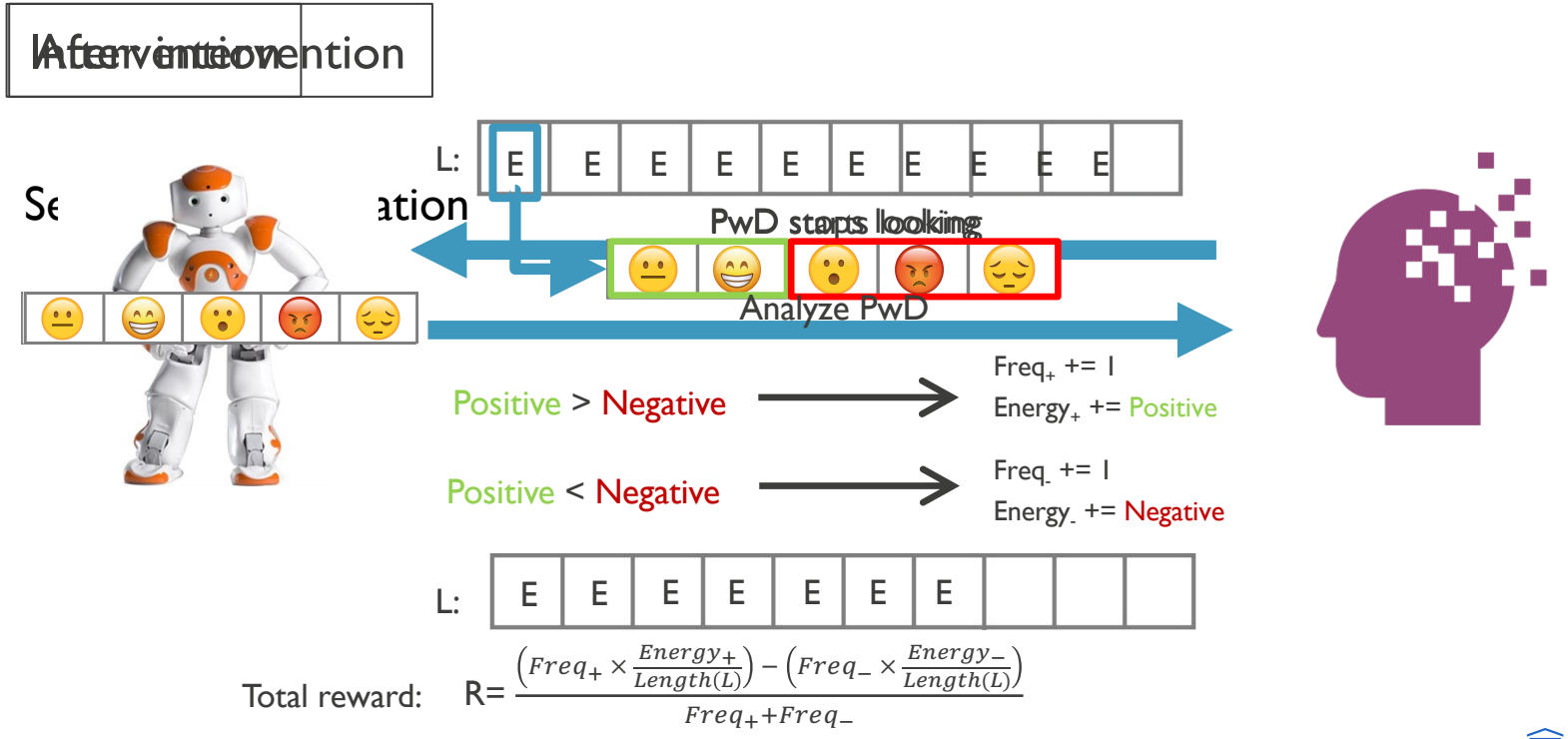
X: Feature vector, representation of the environment
w_a^T: Weight vector, determines the influence of the features

POLICIES

- A balancing paradigm
  - Rand
  - Epsilon-greedy
  - Decaying or a… olicy
  - Upper Confid…
  - Cont…



REWARD SIGNAL

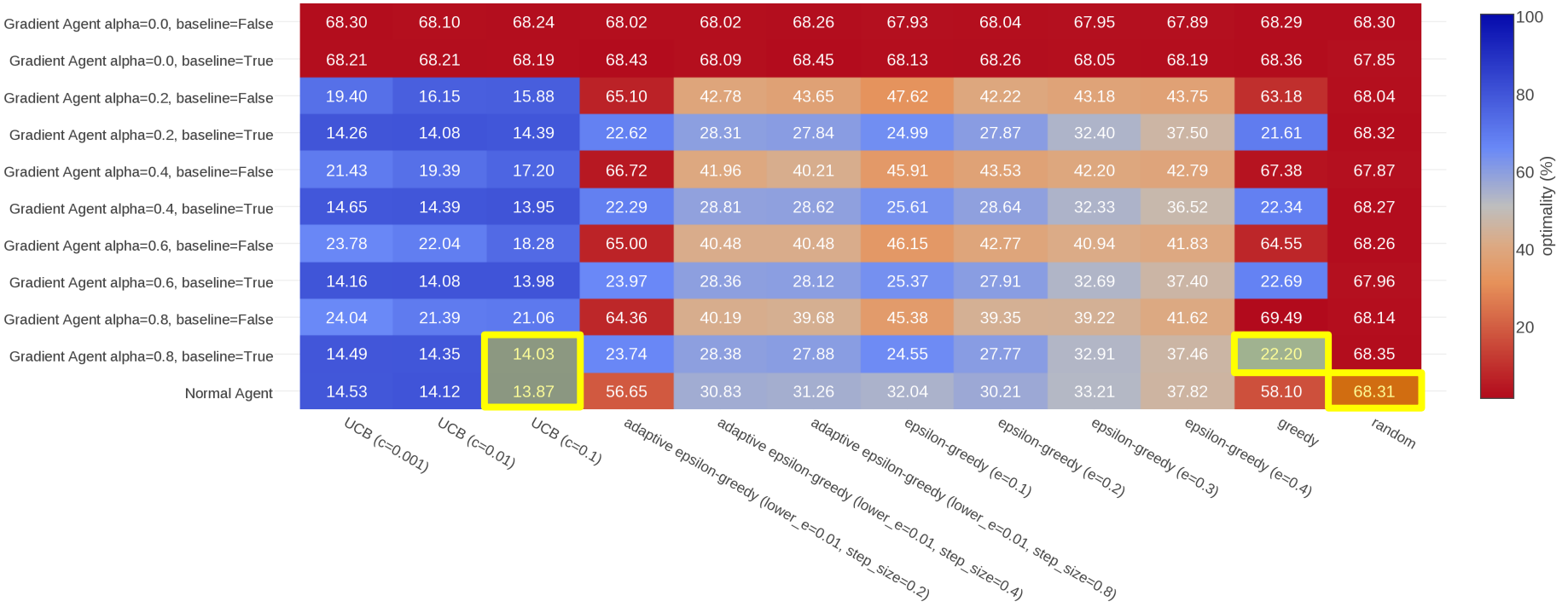


EXPERIMENTS

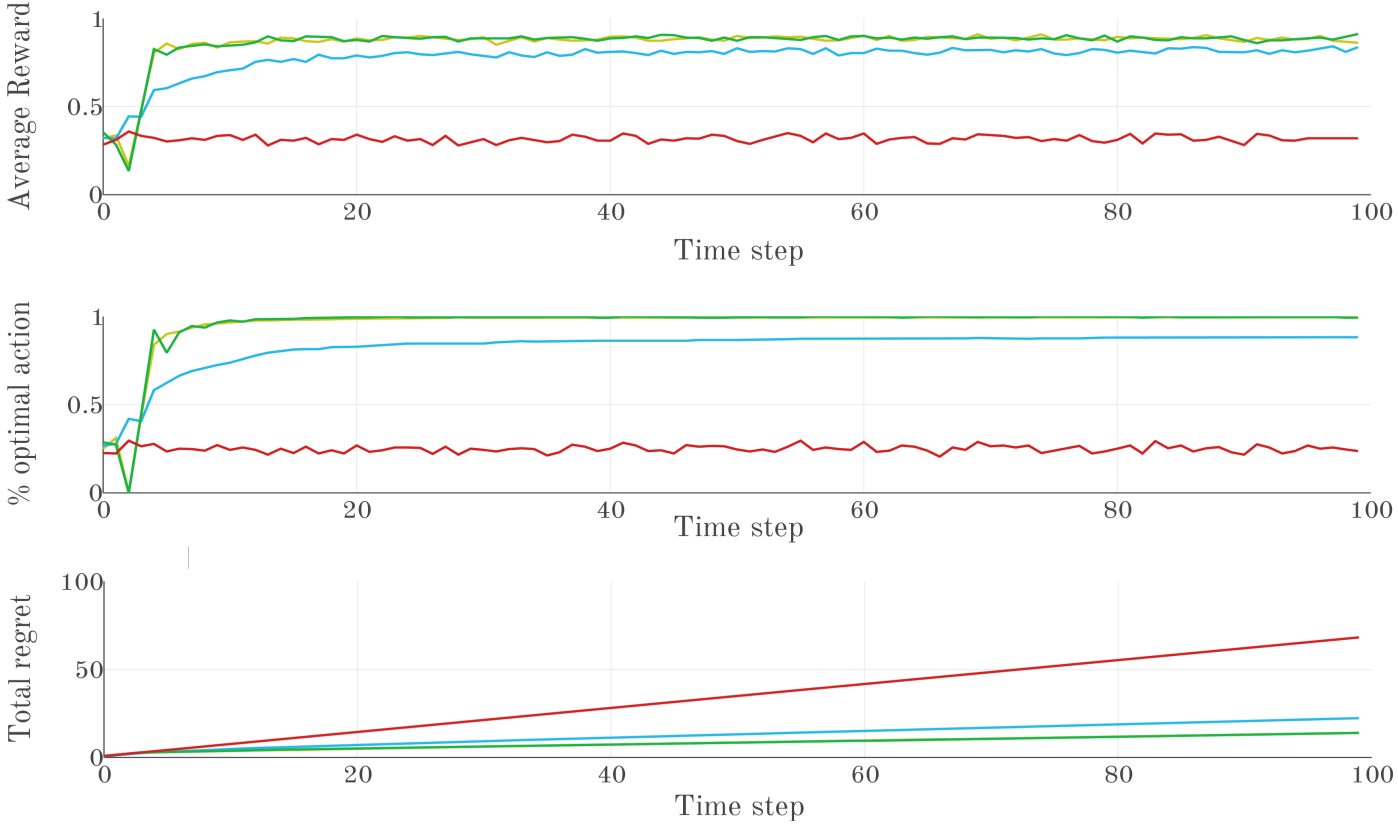
SIMULATION TESTS

- Single patient tests:
  - Policy-agent comparison
  - Comparison for three strategies
  - Goal: search for most promising
- Multi-patients tests:
  - Policy-agent comparison
  - Comparison for three strategies
  - Change in behavioural pattern
  - Goal: Compare personalised learning approach with a global learning one

SINGLE PATIENT AGENT-POLICY COMPARISON



AGENT-POLICY OPTIMISTIC COMPARISON

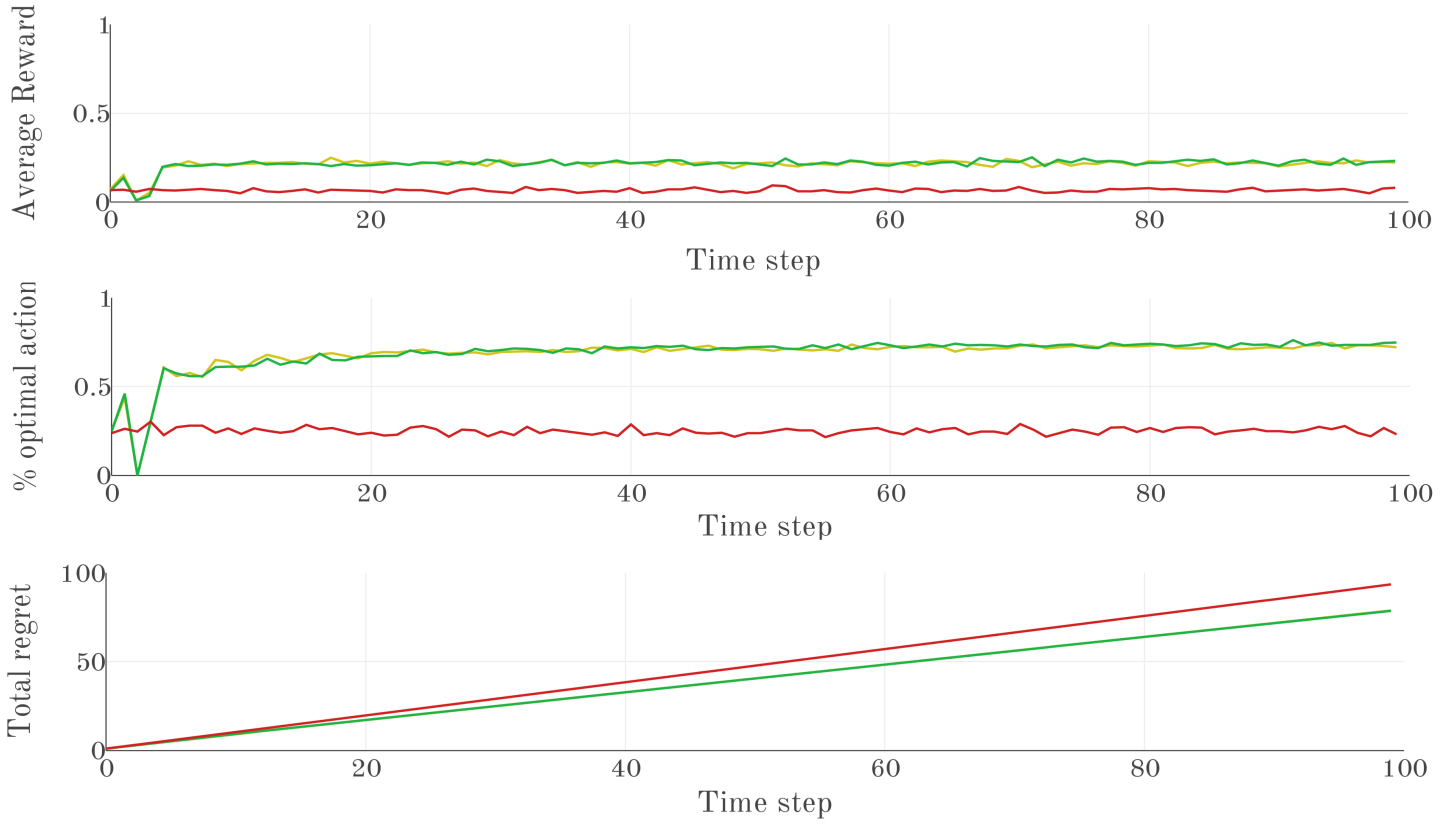


imec

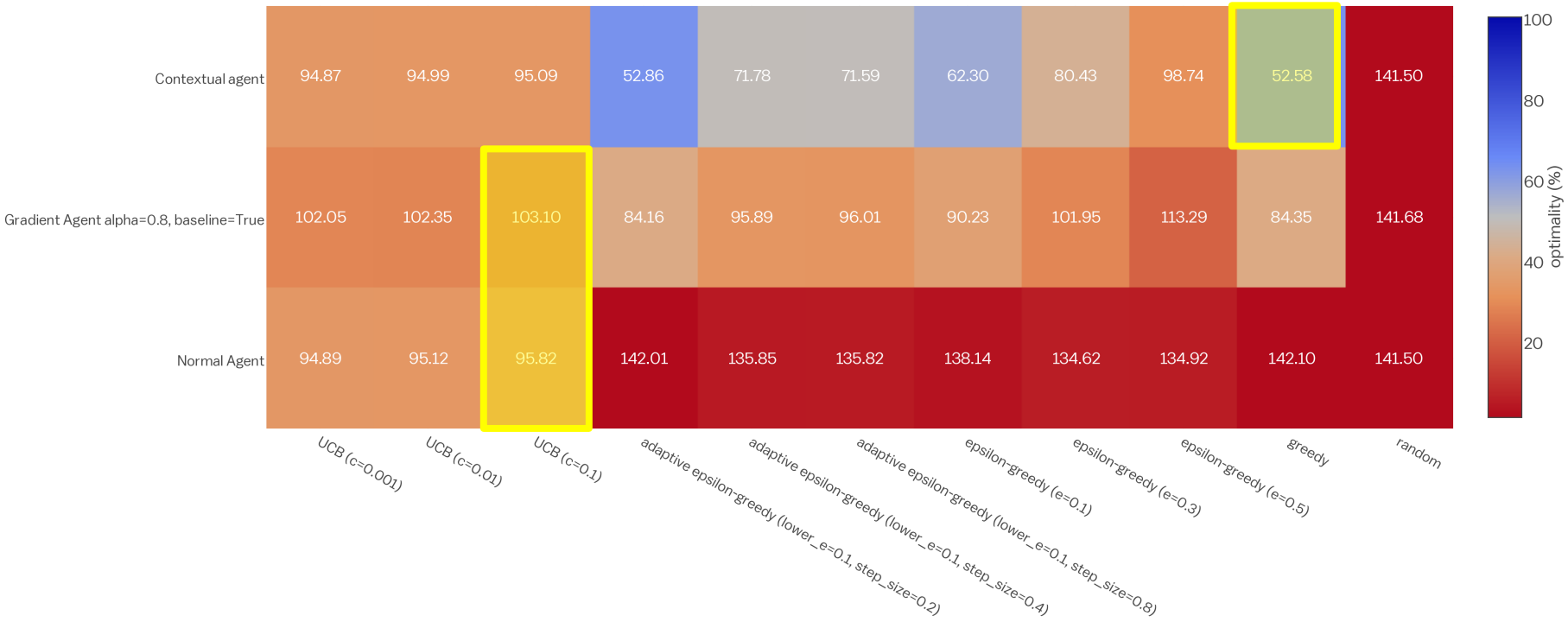
HENT NIVERSITY

Universiteit Antwerpen

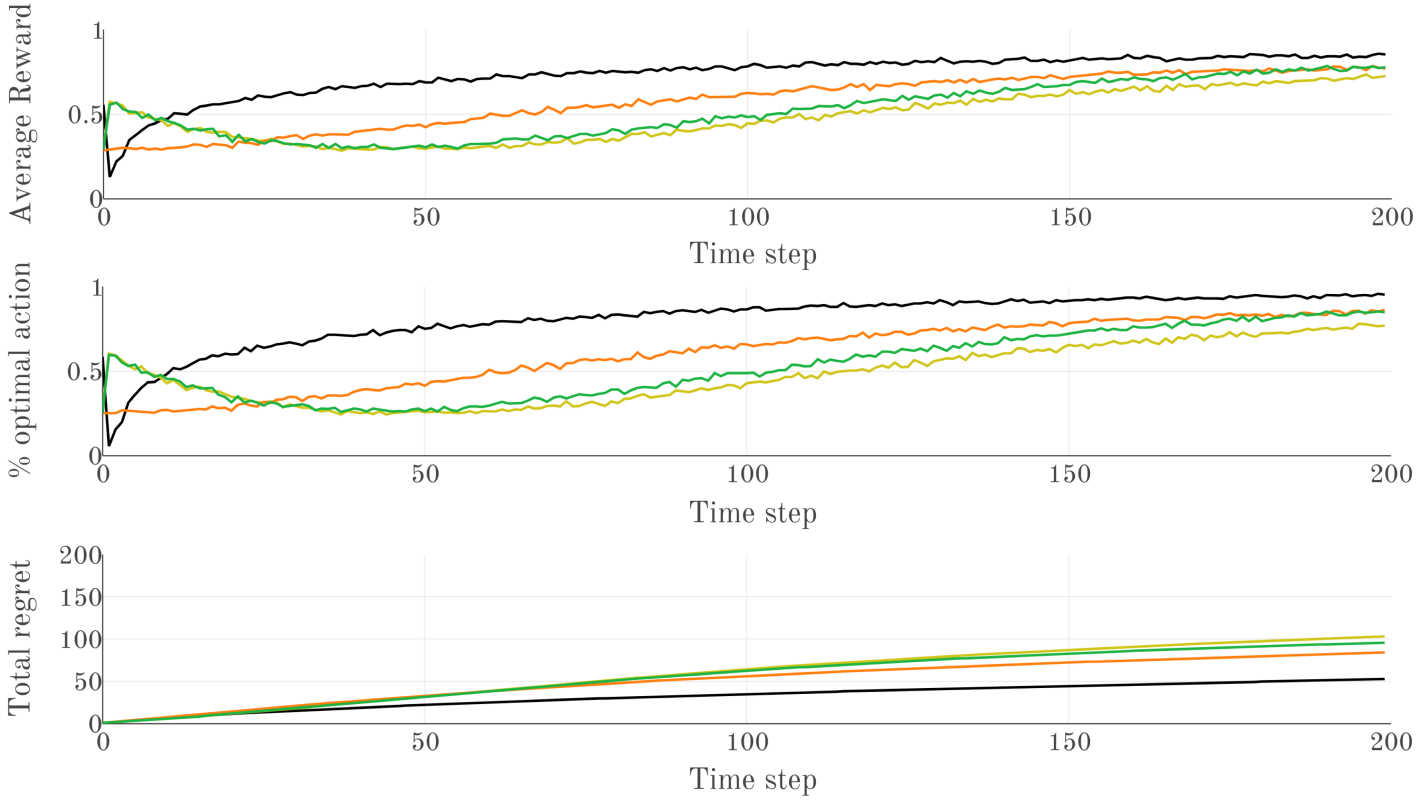
AGENT-POLICY WORST-CASE COMPARISON



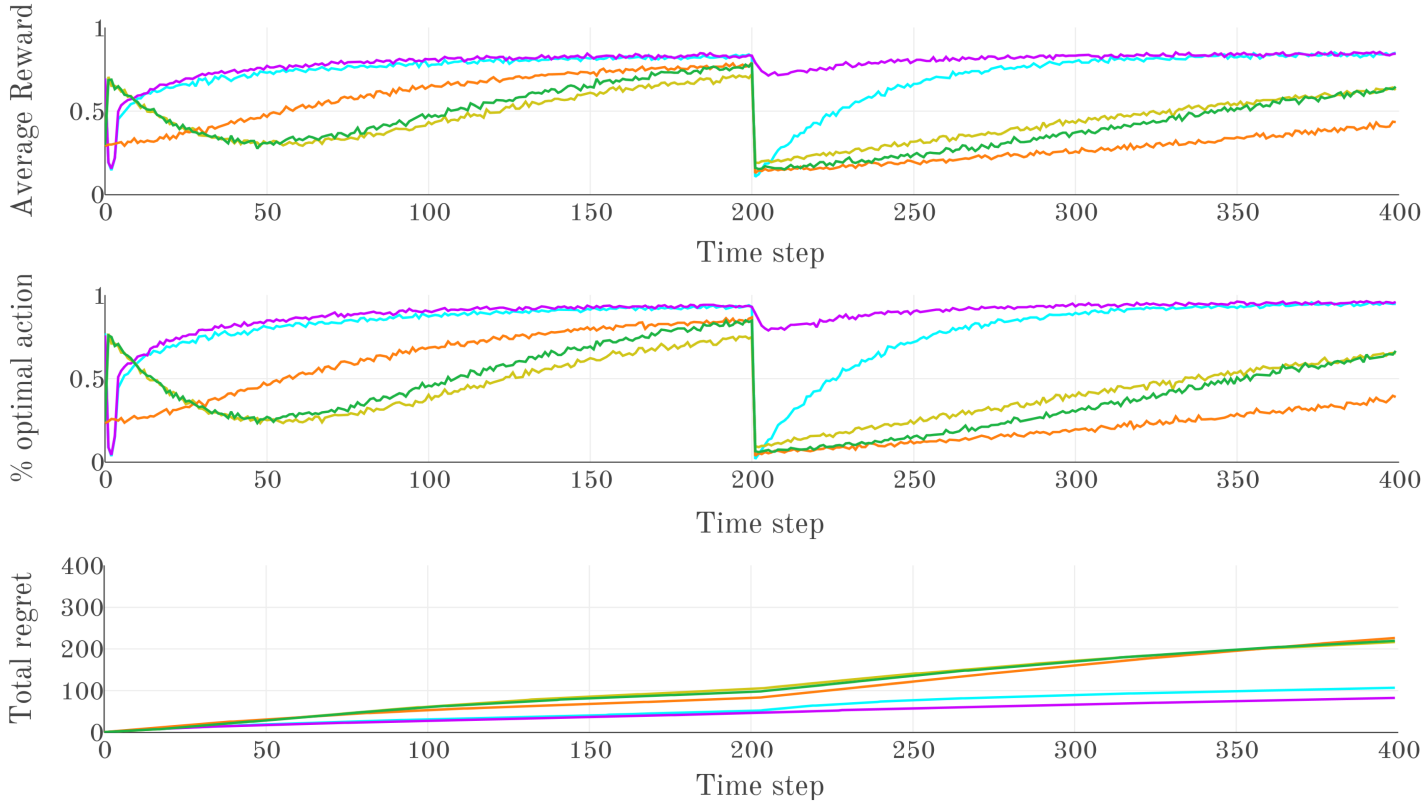
MULTI PATIENT AGENT-POLICY COMPARISON



AGENT-POLICY OPTIMISTIC COMPARISON



AGENT-POLICY OPTIMISTIC COMPARISON



imec

GHENT UNIVERSITY

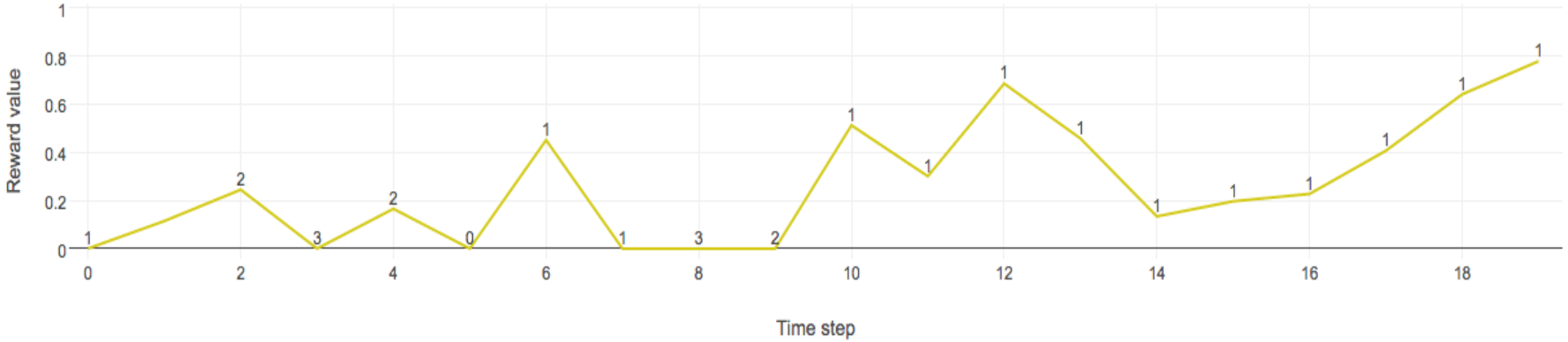
Universiteit Antwerpen

REAL ROBOTIC TESTS

– Task:

action	reaction
0	Show a neutral face
1	Smile or laugh
2	Look scared or bored
3	Be angry/do not look at the robot

action	neutral	happy	surprised	angry	sad
0	0.10785714	0.00642857	0.01333333	0.13785714	0.2345238
1	0.05774646	0.33179213	0.10022848	0.07839935	0.35491048
2	0.12407637	0.06319917	0.03101135	0.0945098	0.35386996
3	0.	0.	0.	0.	0.



CONCLUSION

CONCLUSION

- Simple bandits are promising research area to enable personalized care
- Adaptation to changes in behavioural patterns

  → Contextual agents are able to handle changes in behaviour
- Admission of new resident

  → Data from other patients benefit learning rate

Future research: Advanced tests with real PwD, Additional reward development