Royal College of Anaesthetists

WINTER SYMPOSIUM
3–4 December 2020

Automating anaesthetic delivery with deep reinforcement learning

Chair: Dr Jamie Strachan

Professor Emery Brown
Dr Marcus Badgeley
Dr Gabriel Schamberg

Automating Anaesthetic Delivery with Deep Reinforcement Learning

December 4, 2020

Gabriel Schamberg, Marcus Badgeley, Emery Brown

Neuroscience Statistics Research Laboratory

MIT

The Picower Institute for Learning and Memory

Outline

Background

- Closed loop anaesthetic delivery & reinforcement learning

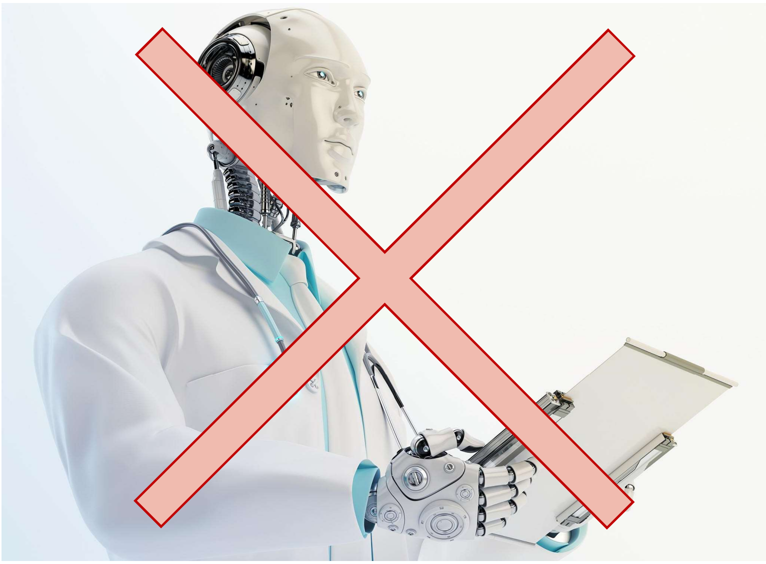
Methods

- Deep RL paradigm for controlling level of unconsciousness

Results

- Simulation study

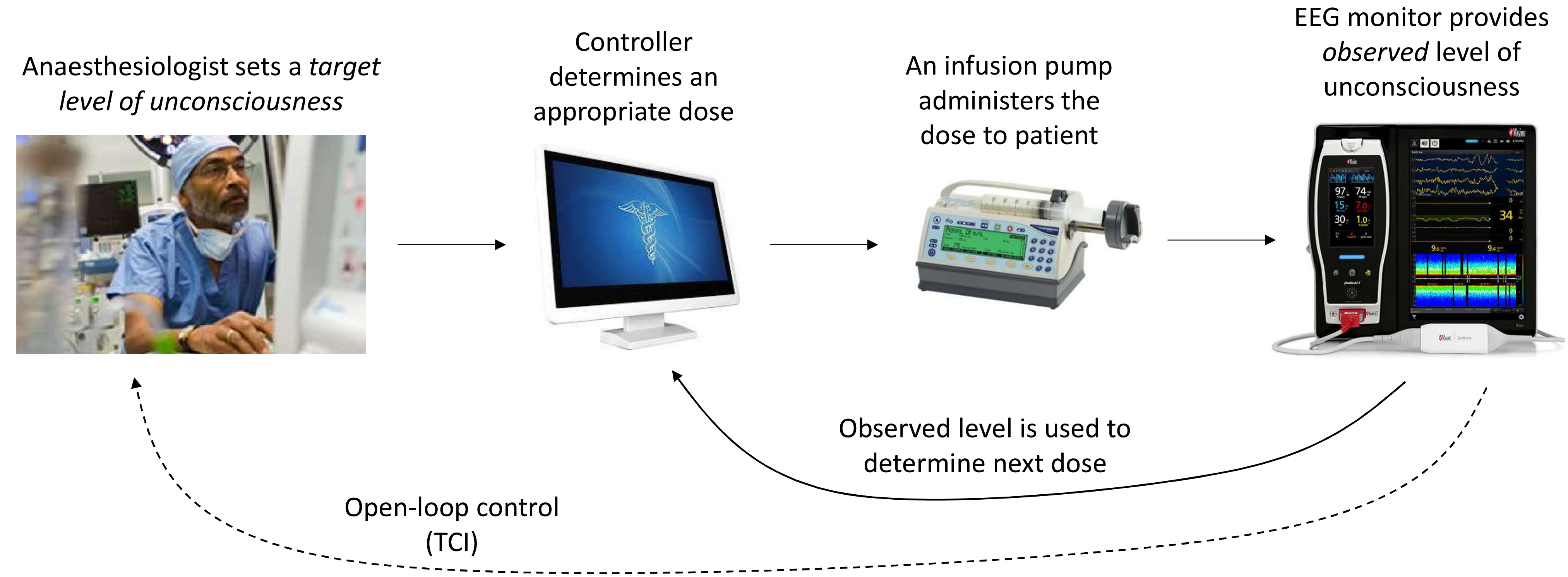
Autopilot for Anaesthesia



THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

Image Sources:
https://www.thetimes.co.uk/article/is-there-a-robot-doctor-in-the-house-g578cpqb2
https://www.cntraveler.com/story/how-autopilot-on-planes-works

Closed Loop Anesthetic Delivery (CLAD)



THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

Image Sources:
https://datalux.com/products/medical-and-enterprise-grade-all-in-one-computers-for-healthcare/
http://www.medfusionpump.com/tour/
https://www.masimo.com/products/continuous/root/root-sedline/

CLAD – What's Been Done?

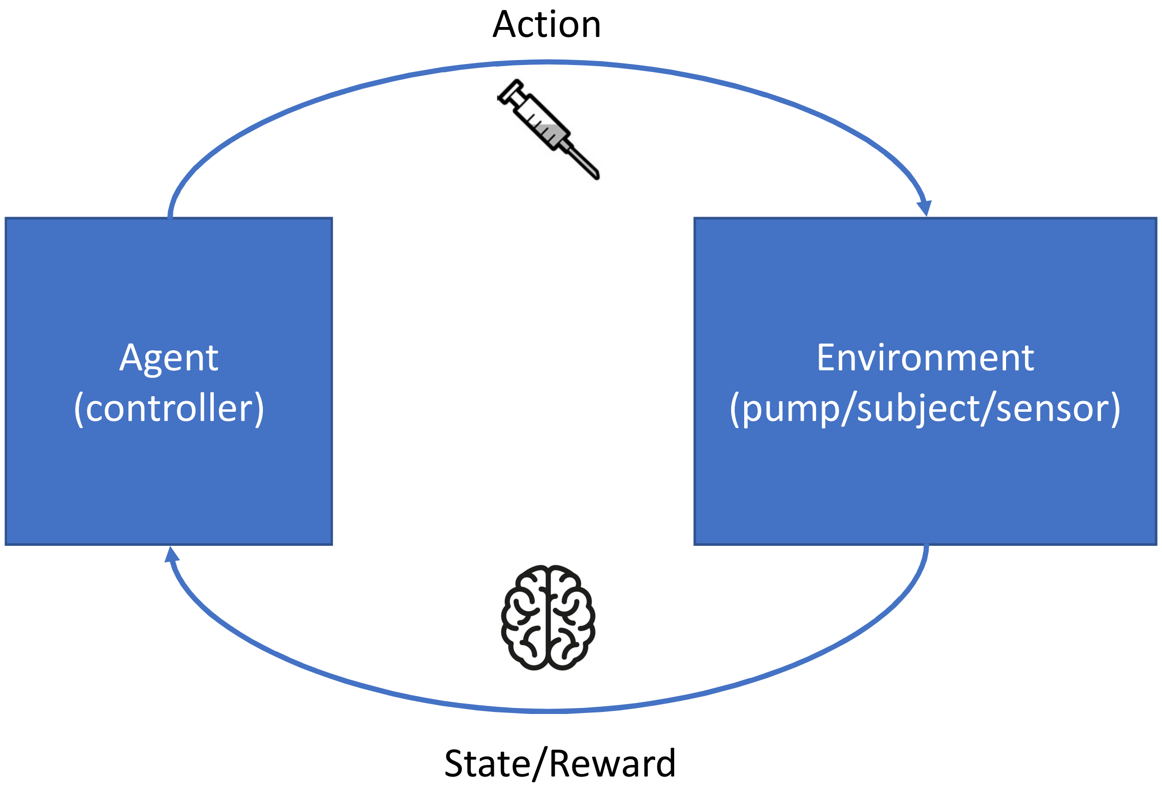
Paper	Subject	Controller	Level of Unconsciousness
Absalom (2002)	Human	PID	BIS
Dumont (2009)	Simulation	Robust PID/CRONE	WAV_{CNS}
Shanechi (2013)	Rodent	LQR	BSP
Moore (2014)	Human	Tabular RL	BIS
Many More…			

Reinforcement Learning

The agent observes a **state**

Based on the state, the agent uses its **policy** to determine an **action**

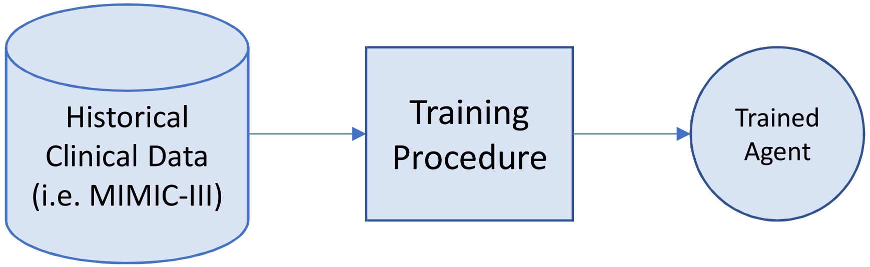
Over time, the agent receives **rewards**, which are used to update its policy



Reinforcement Learning – Two Approaches

Off-Policy*:*

The agent "watches and learns"



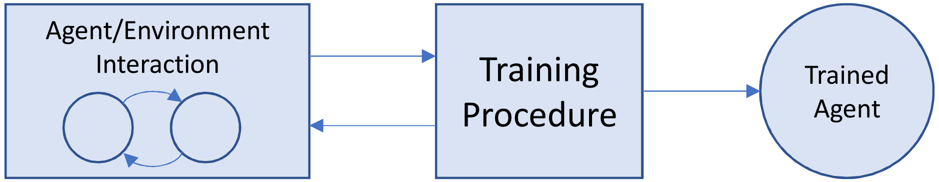
- Agent sees real patient responses
- Agent learns to reflect anaesthesiologist's actions

- Data may not contain enough bad or ideal behavior
- For example, infrequent titration is suboptimal

On-Policy*:*

The agent "learns by doing"

Focus of this talk



- Agent can fully explore action space
- The actions will respond to varying reward functions

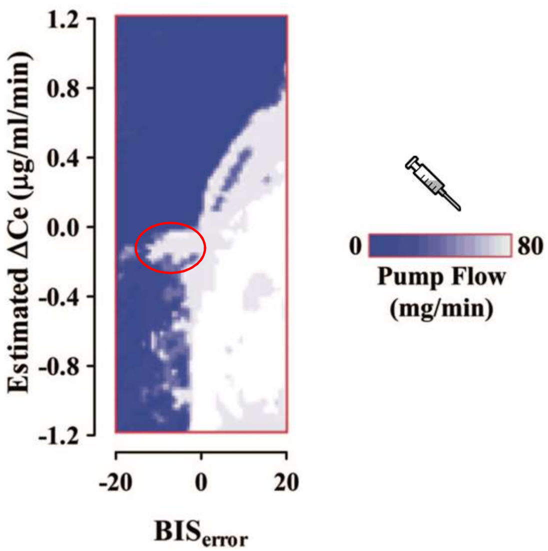
- Agent cannot act on humans → requires simulations

Tabular RL Policies

A **<u>policy</u>** maps a **<u>state</u>** to an **<u>action</u>**

Tabular RL learns which discrete action should be taken for every *discrete state*

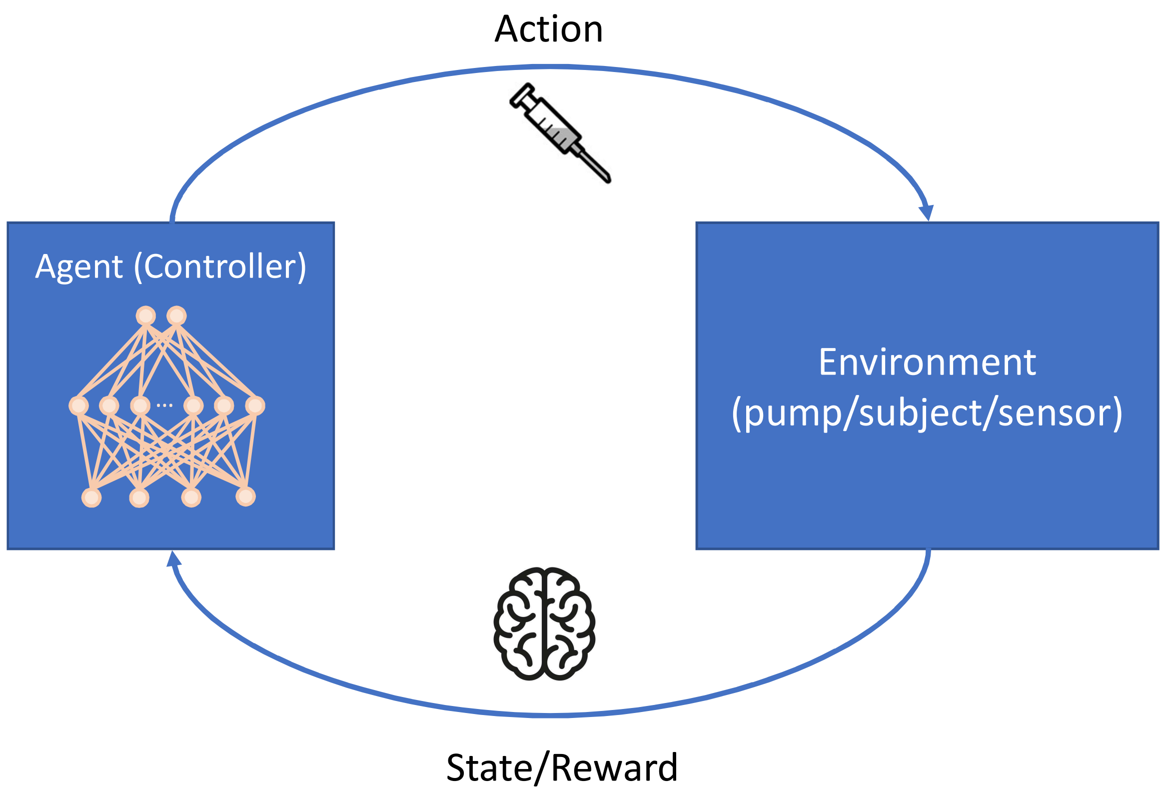
- Ignores the continuous relationship between state and action
- Results in patchy discontinuities
- Dimensionality of the table scales exponentially with the state space



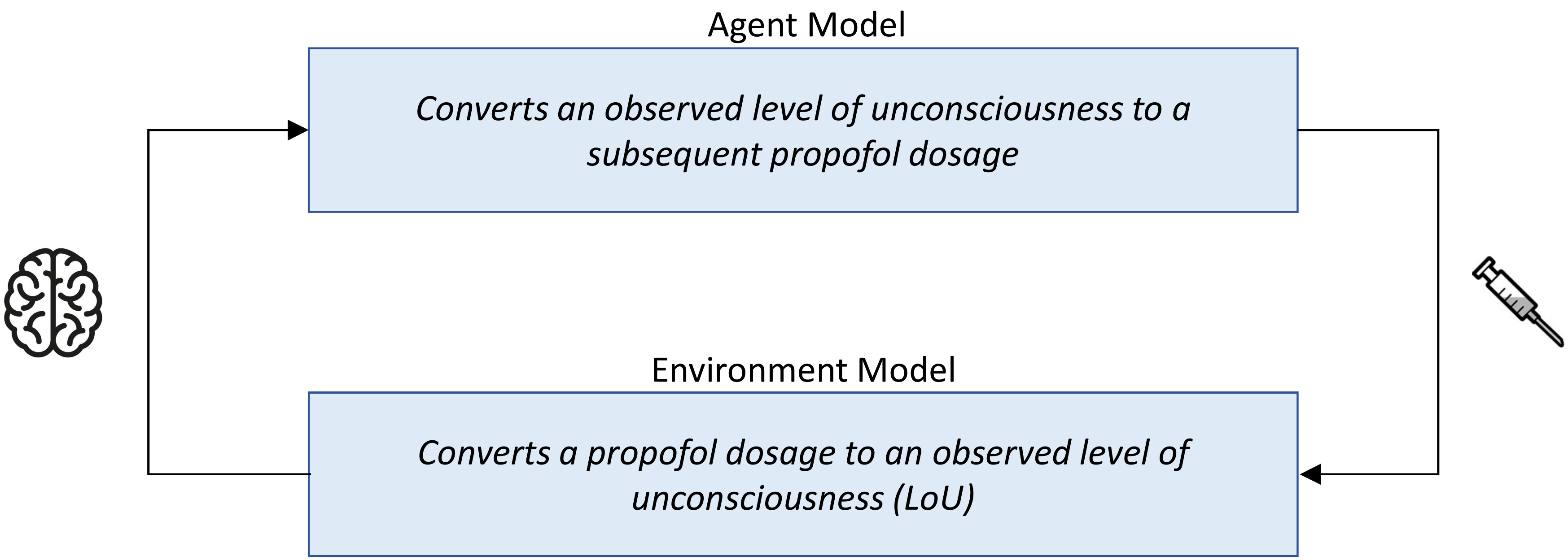
From Moore et al. (2011)

<u>Deep</u> Reinforcement Learning

The policy is a
<u>deep neural network</u>



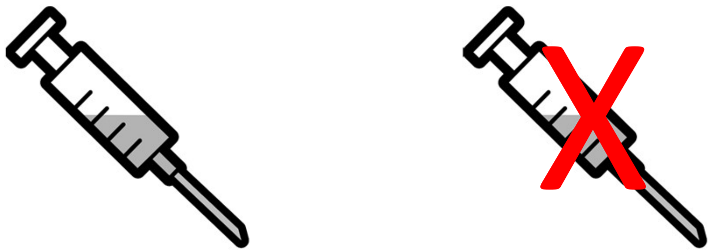
Deep RL for CLAD



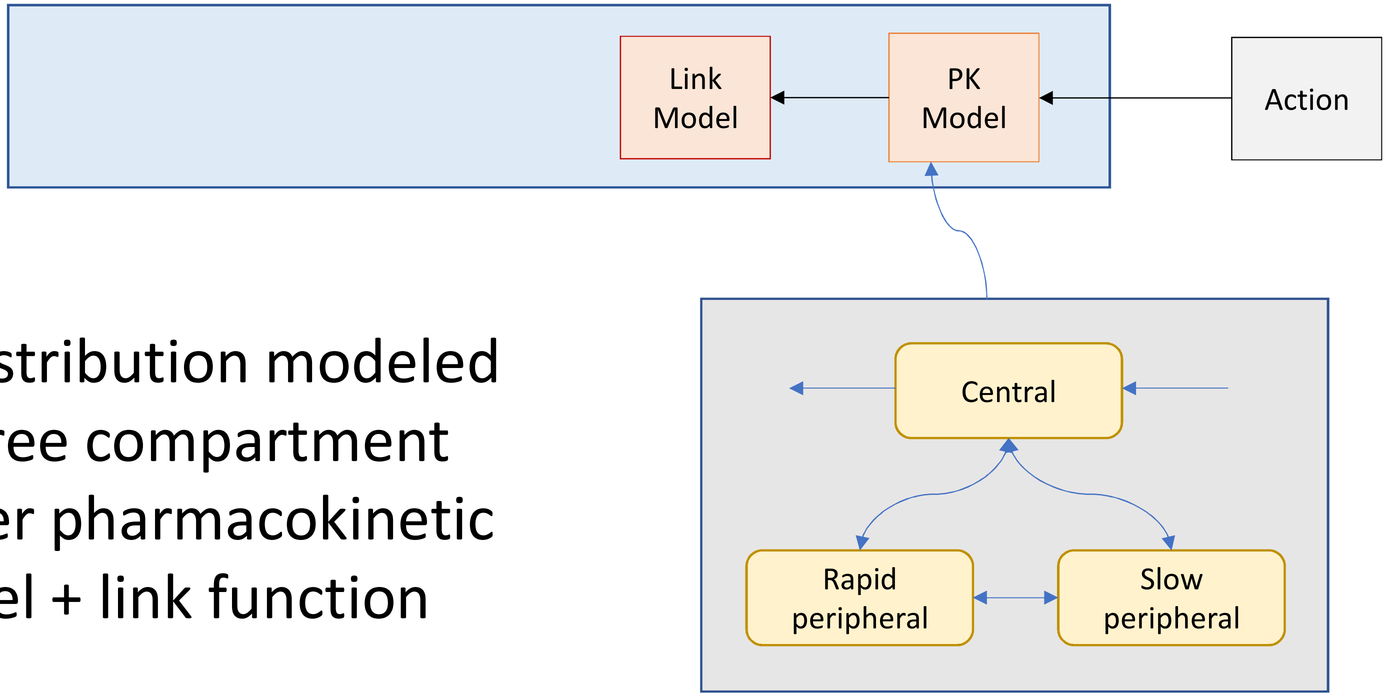
THE PICOWER INSTITUTE
FOR LEARNING AND MEMORY

The Environment Model

5 seconds of
1.67 mg/s or **0 mg/s**



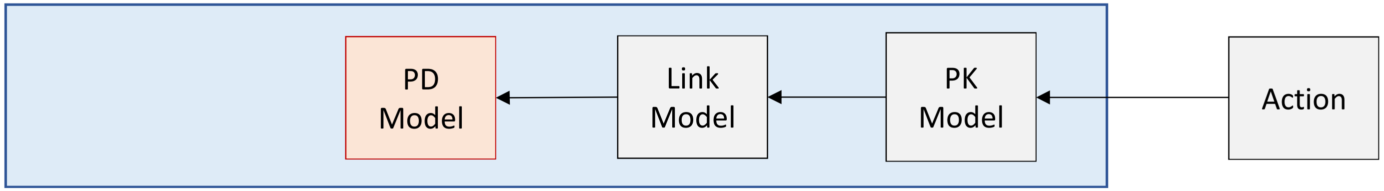
The Environment Model



Drug distribution modeled by three compartment Schnider pharmacokinetic model + link function

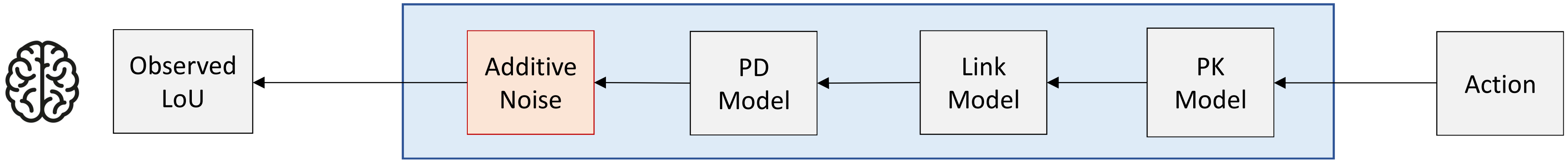
THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

The Environment Model

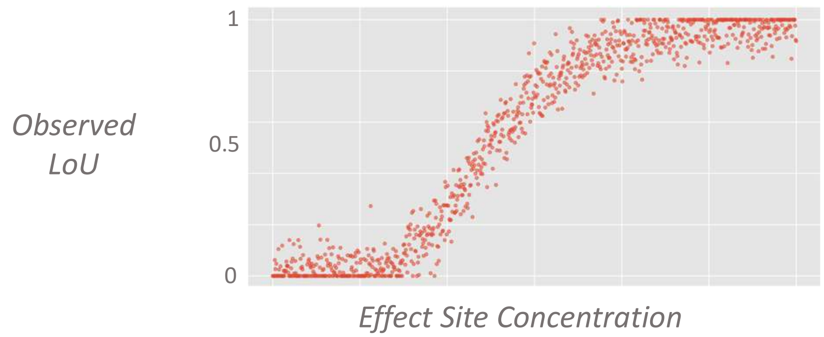


Drug effect modeled by nonlinear pharmacodynamic model determines level of unconsciousness (LoU)

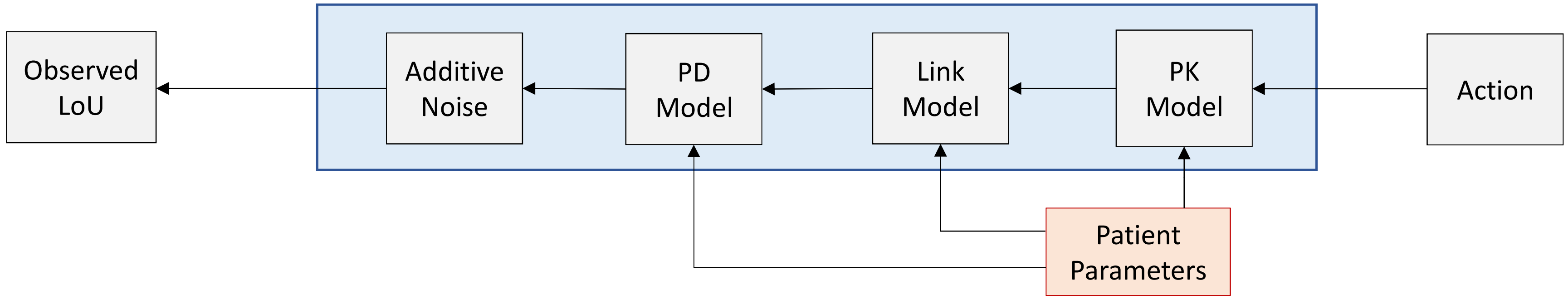
The Environment Model



Observed LoU includes additive noise



The Environment Model

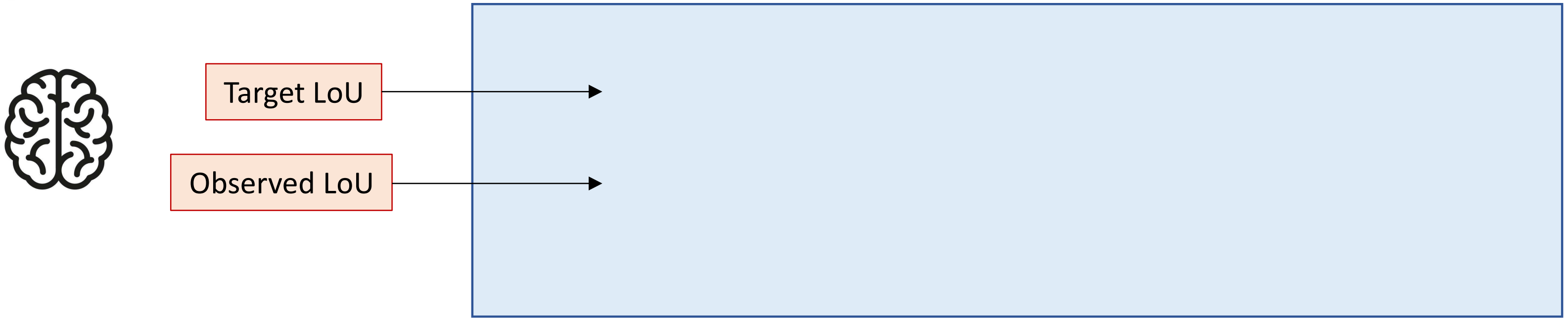


patient variability model

Sub-model	Parameter	Units	Generic	Minimum	Maximum
PK	Height	cm	170	160	190
PK	Weight	kg	70	50	100
PK	Age	yr	30	18	90
Link	k_{e0}	min^{-1}	0.17	0.128	0.213
PD	γ	-	5	5	9
PD	C	-	2.5	2	6

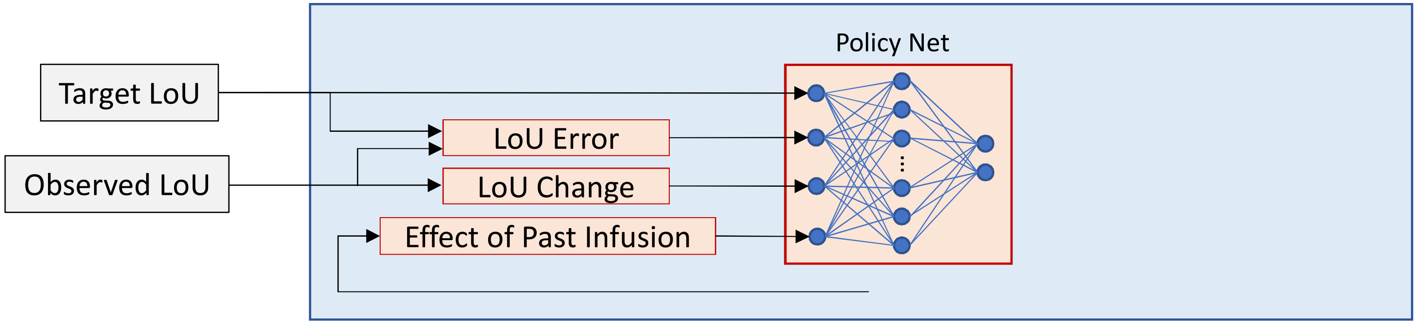
THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

The Agent Model



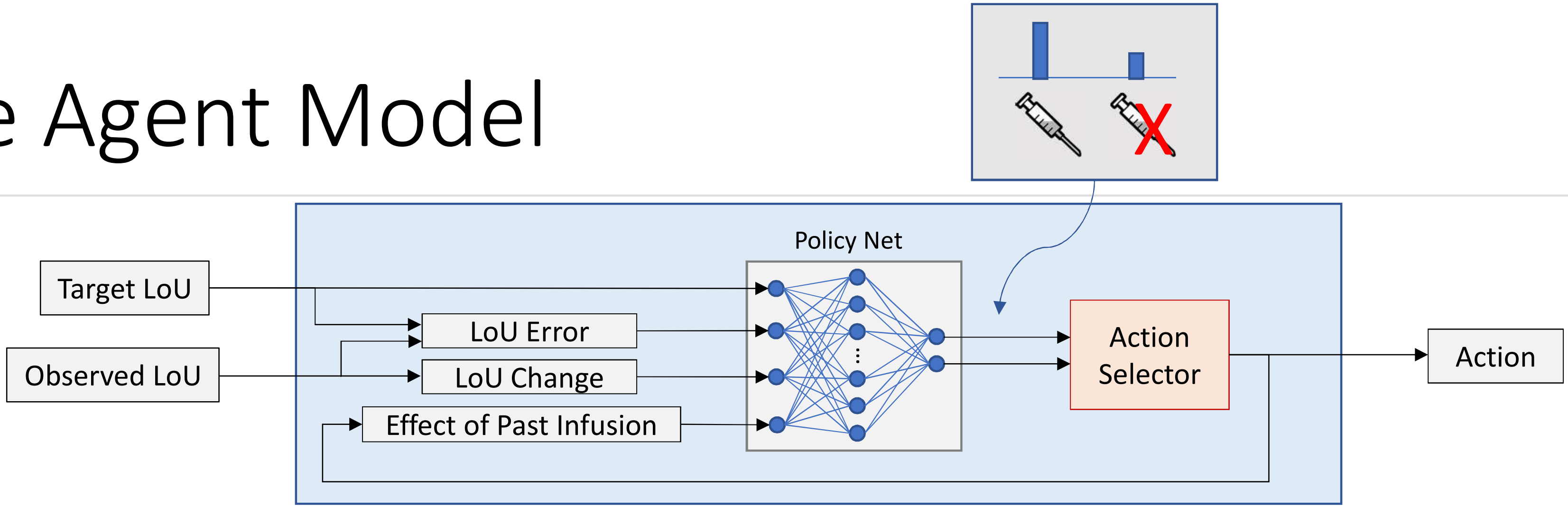
Agent receives observed
and target LoUs

The Agent Model



The policy assigns probabilities to actions based on a 4 dimensional observation

The Agent Model



Agent uses one of three strategies to select an action

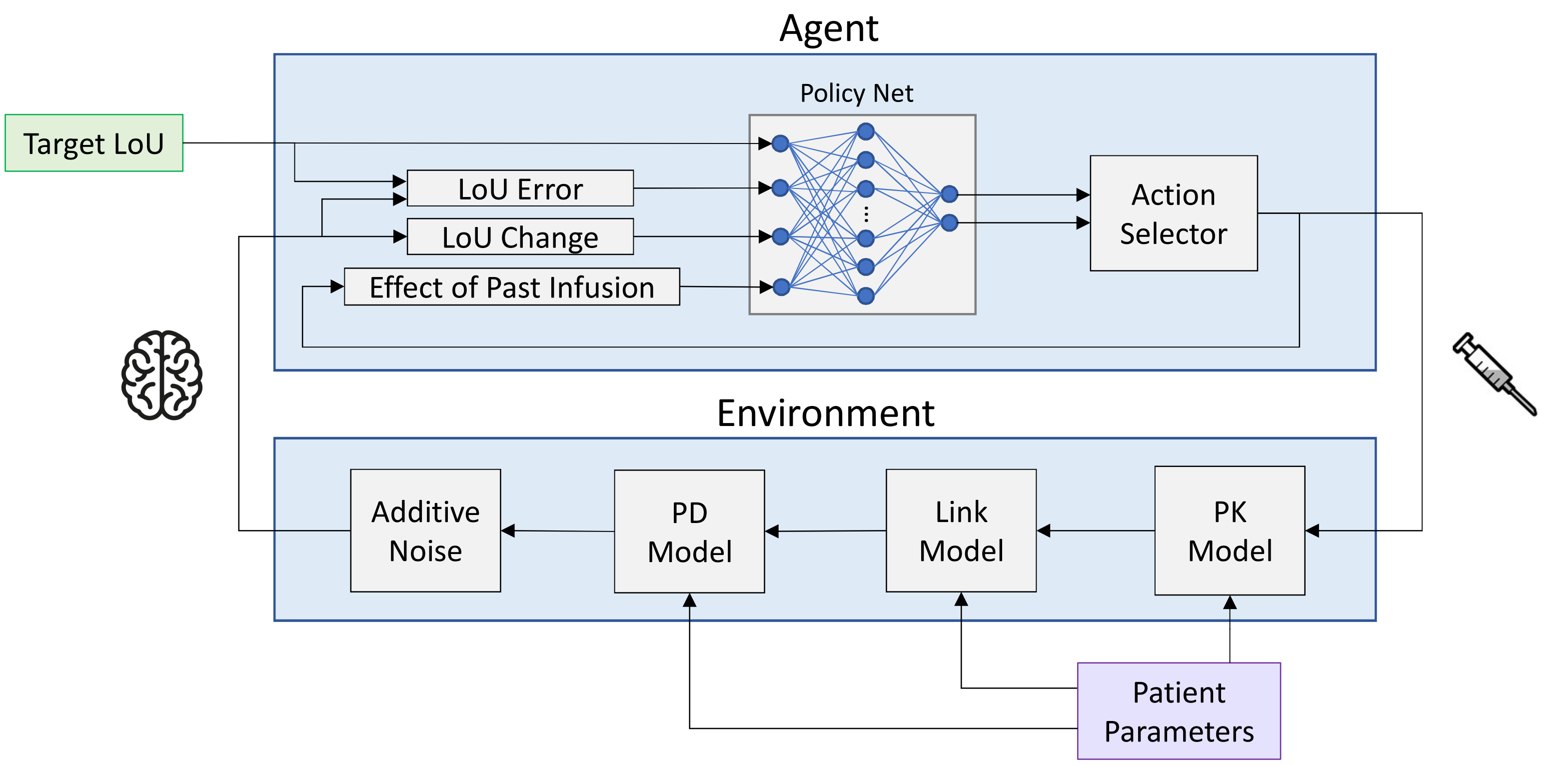
__Stochastic__
Picks action randomly according policy output

__Deterministic__
Pick action with highest probability

__Continuous__
Multiply action probability by max dose

THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

Complete Simulation Model

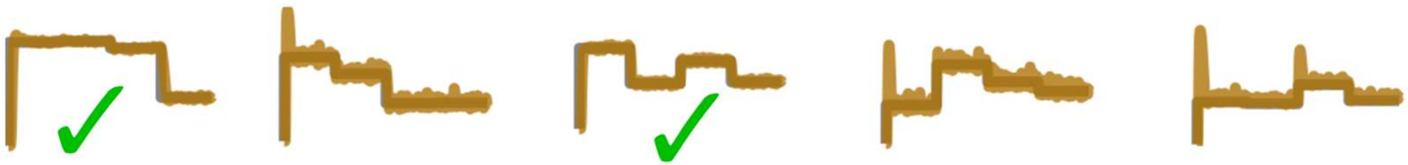


THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

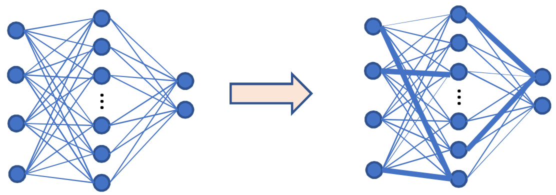
Training the Agent

Agent acts with randomness in a series of “episodes”

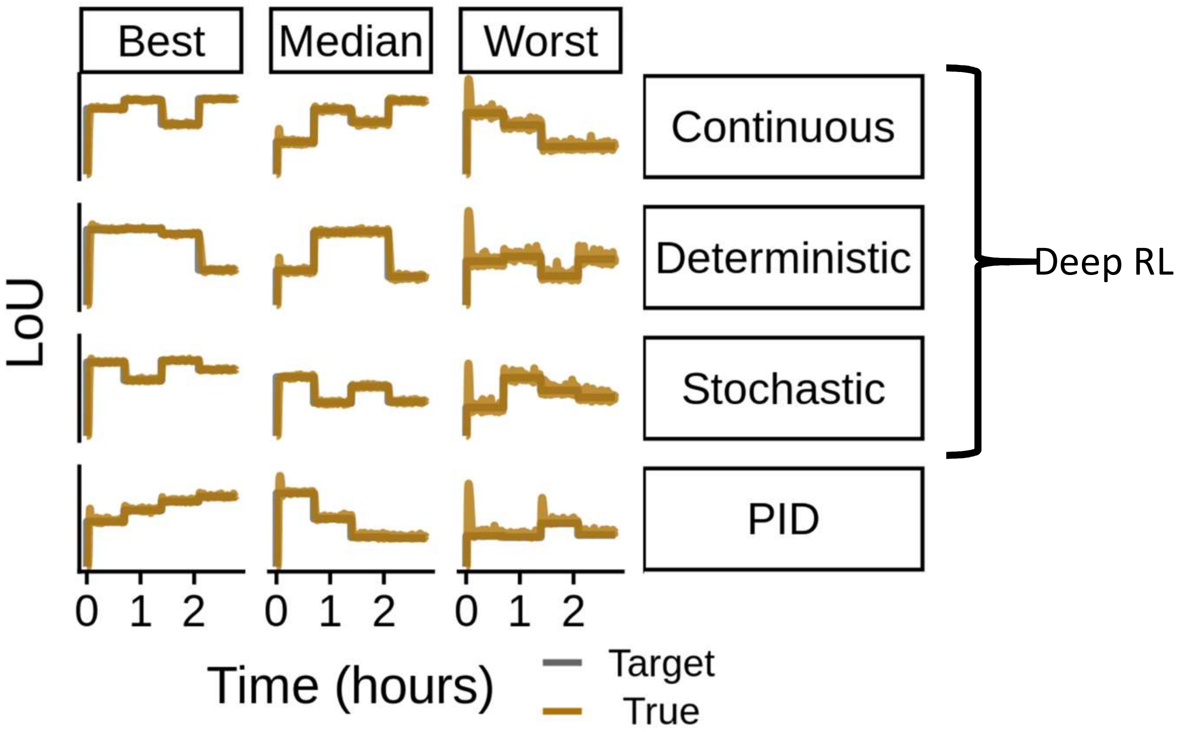
Identify which episodes were best



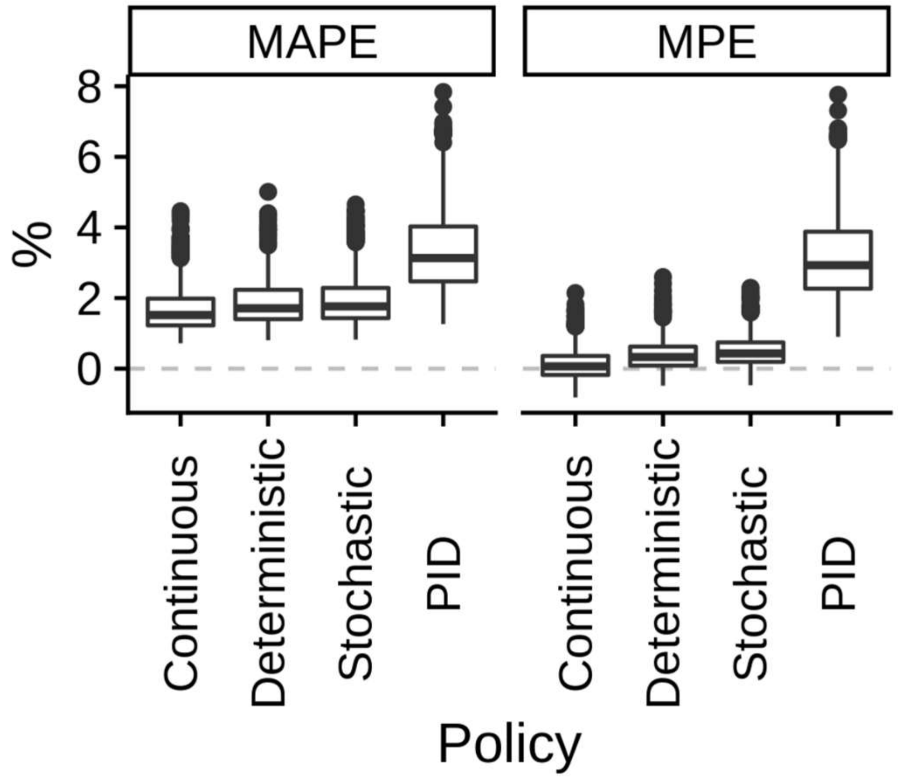
Nudge policy to encourage repeating best actions



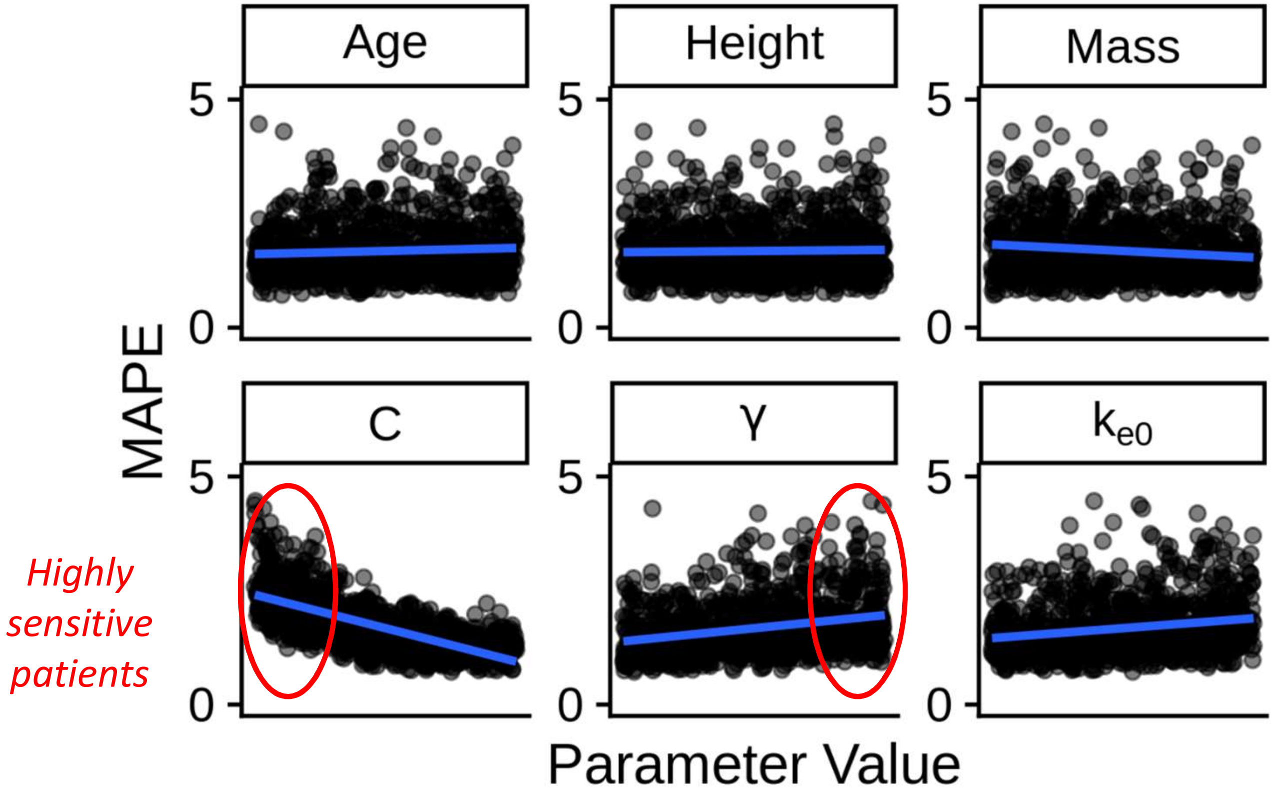
Deep RL Works



Deep RL Outperforms PID

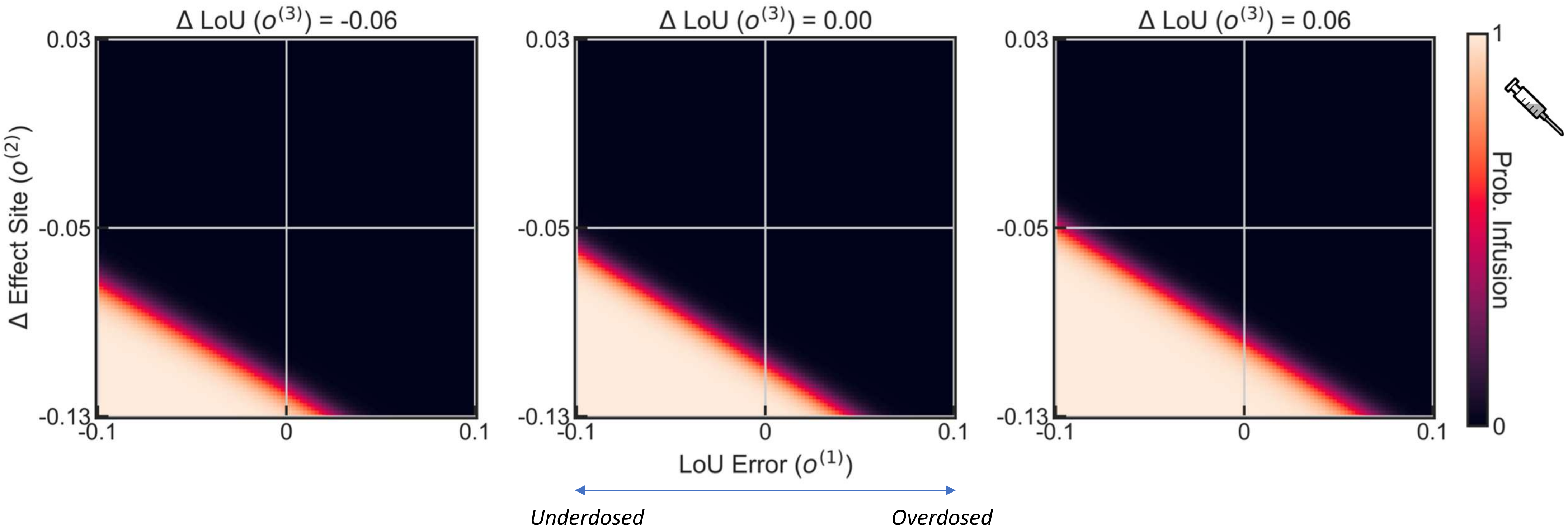


Deep RL is Robust to Patient Variability



THE PICOWER INSTITUTE
FOR LEARNING AND MEMORY

Deep RL is Not a Black Box



THE PICOWER INSTITUTE FOR LEARNING AND MEMORY

Key Conclusions

- Deep RL outperforms PID controller in simulations
- Deep RL resolves issues with tabular RL
  - Enables consideration of other patient variables
  - Reflects continuous state/action relationships
- Resulting policy is not a black box

Future Work

- Further simulation studies
  - Intra-patient variability in environment
  - Continued feature engineering
  - Off-policy training on retrospective clinical data
- Clinical studies
  - Human-in-the-loop recommender system
  - Animal studies
  - Volunteers?

Acknowledgements

- Marcus Badgeley
- Emery Brown
- Benyamin Meschede-Krasa
- John Abel

BACKUP

A Side by Side Comparison of Control Algorithms

<table>
<tr><th>Classical Model-Free (PID)</th><th>Classical Model-Based (LQR)</th><th>Deep RL</th></tr>
<tr><td>Parameters are tuned using nominal transfer function</td><td>Parameters are derived using nominal patient model</td><td>Parameters are learned using patient model simulations</td></tr>
<tr><td>Does not optimize</td><td>Optimizes a quadratic cost</td><td>Optimizes reward</td></tr>
<tr><td>Linear function of error</td><td>Linear function of state</td><td>Non-Linear function of state</td></tr>
<tr><td colspan="2">Well-established methods for testing stability, robustness, responsiveness</td><td>No formal analysis</td></tr>
<tr><td colspan="3">End product is a deterministic controller</td></tr>
</table>