A Brief Introduction to Reinforcement Learning in Medicine

Matthew Engelhard


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO


NETFLIX
ALPHAGO

Sequential Decision-Making

Make a series of decisions

based on a set of features (state)

to maximize reward over time

Sequential Decision-Making

Make a series of decisions

based on a set of features (state)

to maximize reward over time

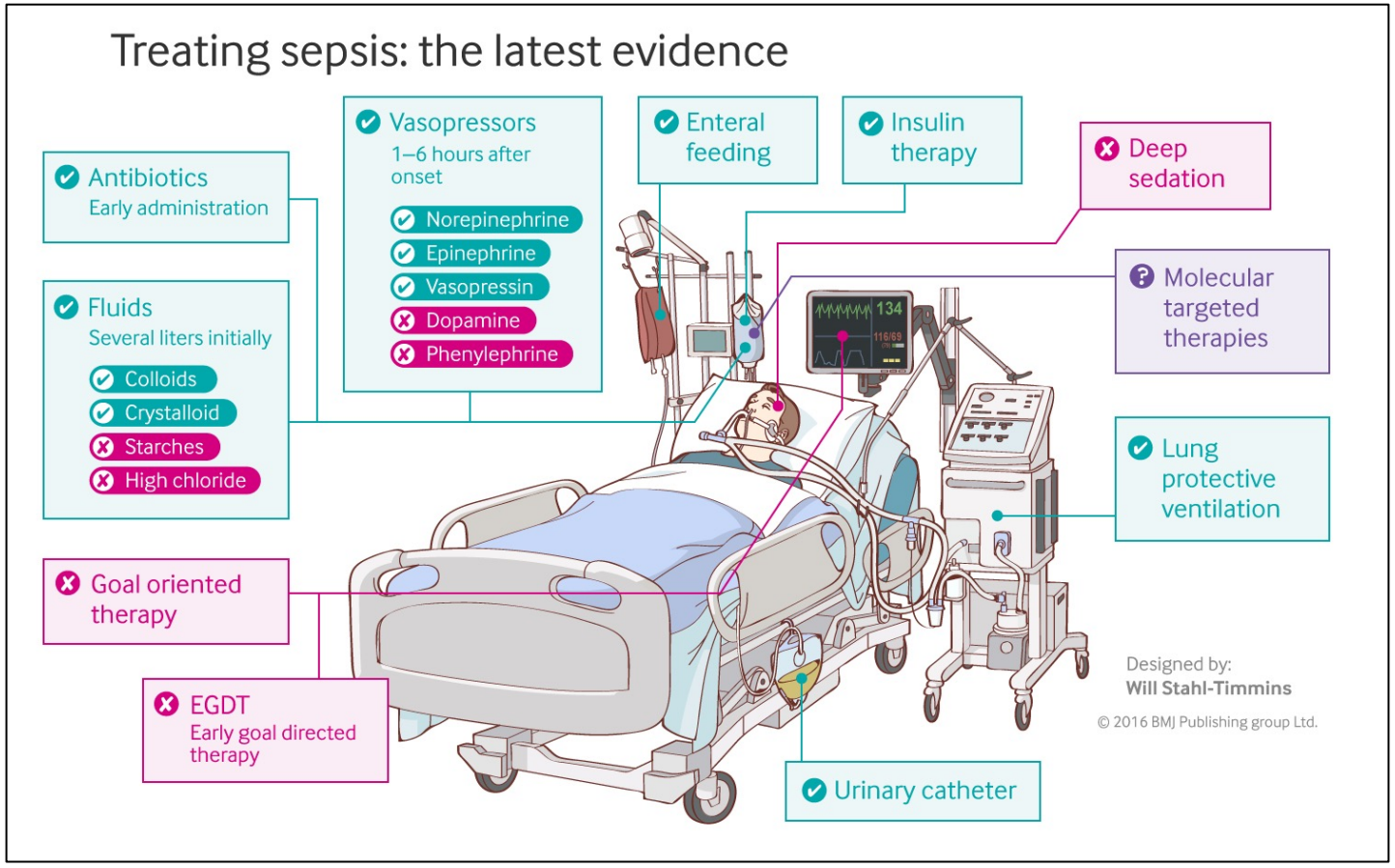
https://www.techradar.com/news/fords-robot-postman-will-deliver-packages-to-your-front-door

Sequential Medical Decision-Making

Make a series of decisions

based on a set of features (state)

to maximize reward over time

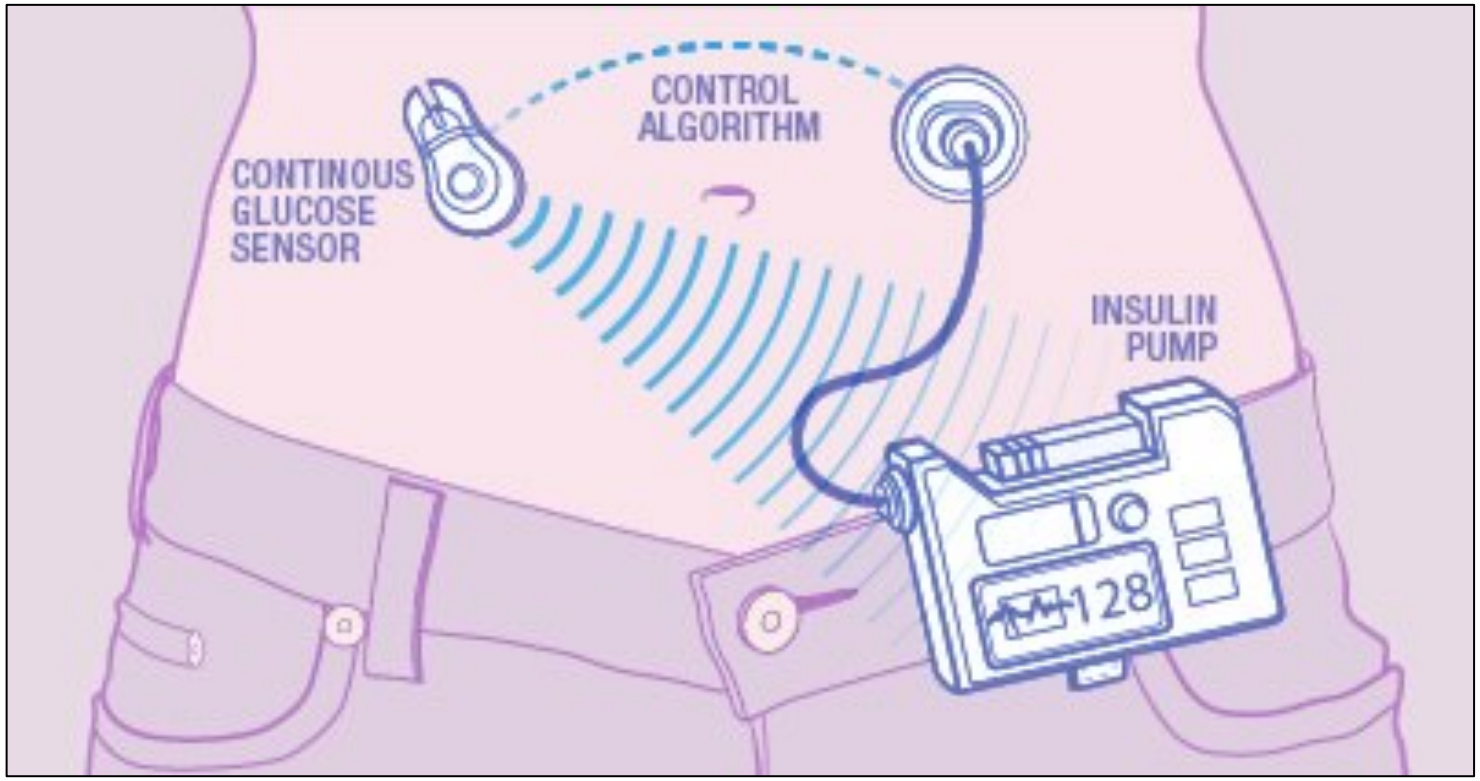


Sequential Medical Decision-Making

Make a series of decisions

based on a set of features (state)

to maximize reward over time

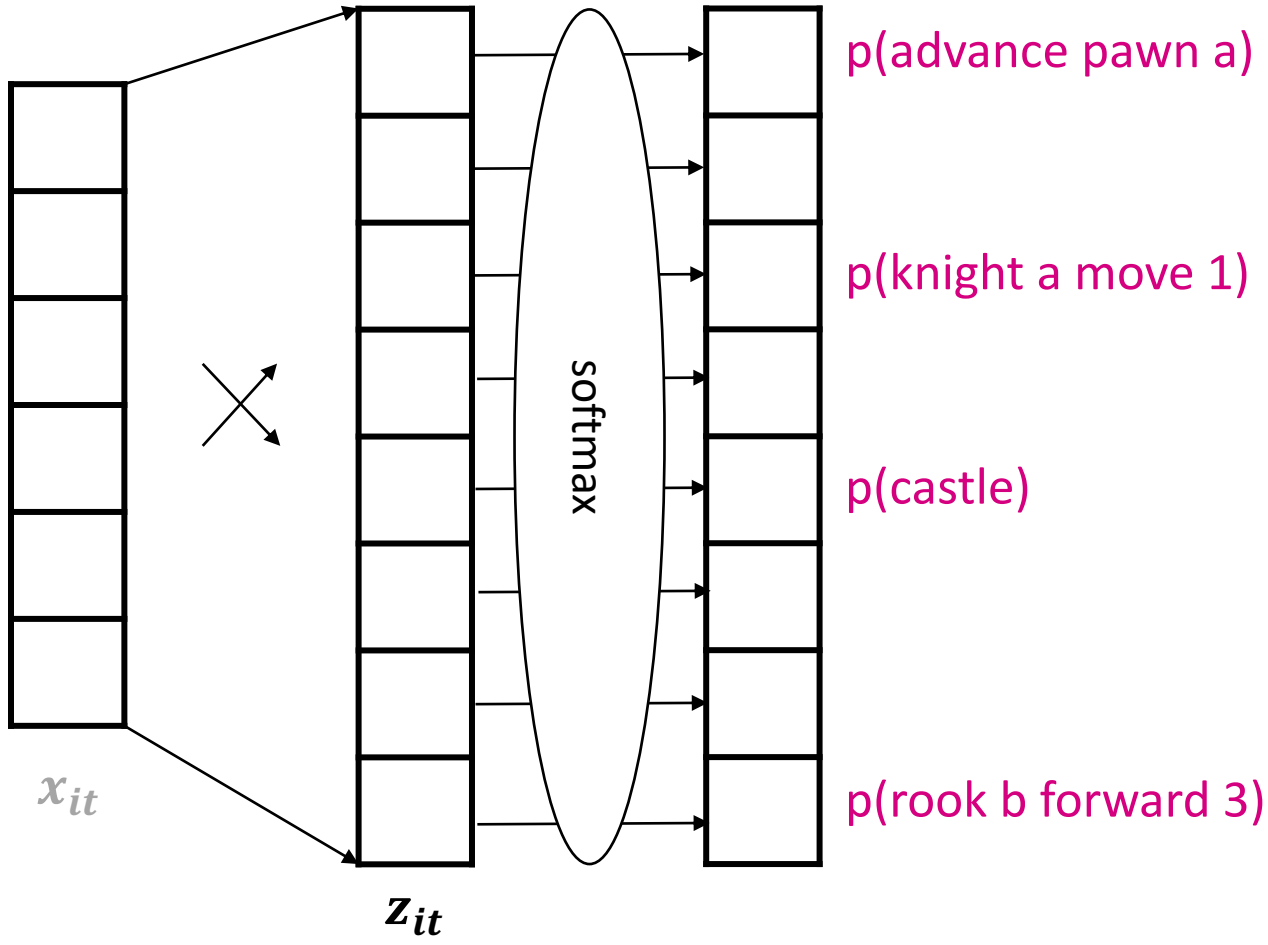


Similarities to models we've discussed

Make a series of decisions

based on a set of features (state)

to maximize reward over time

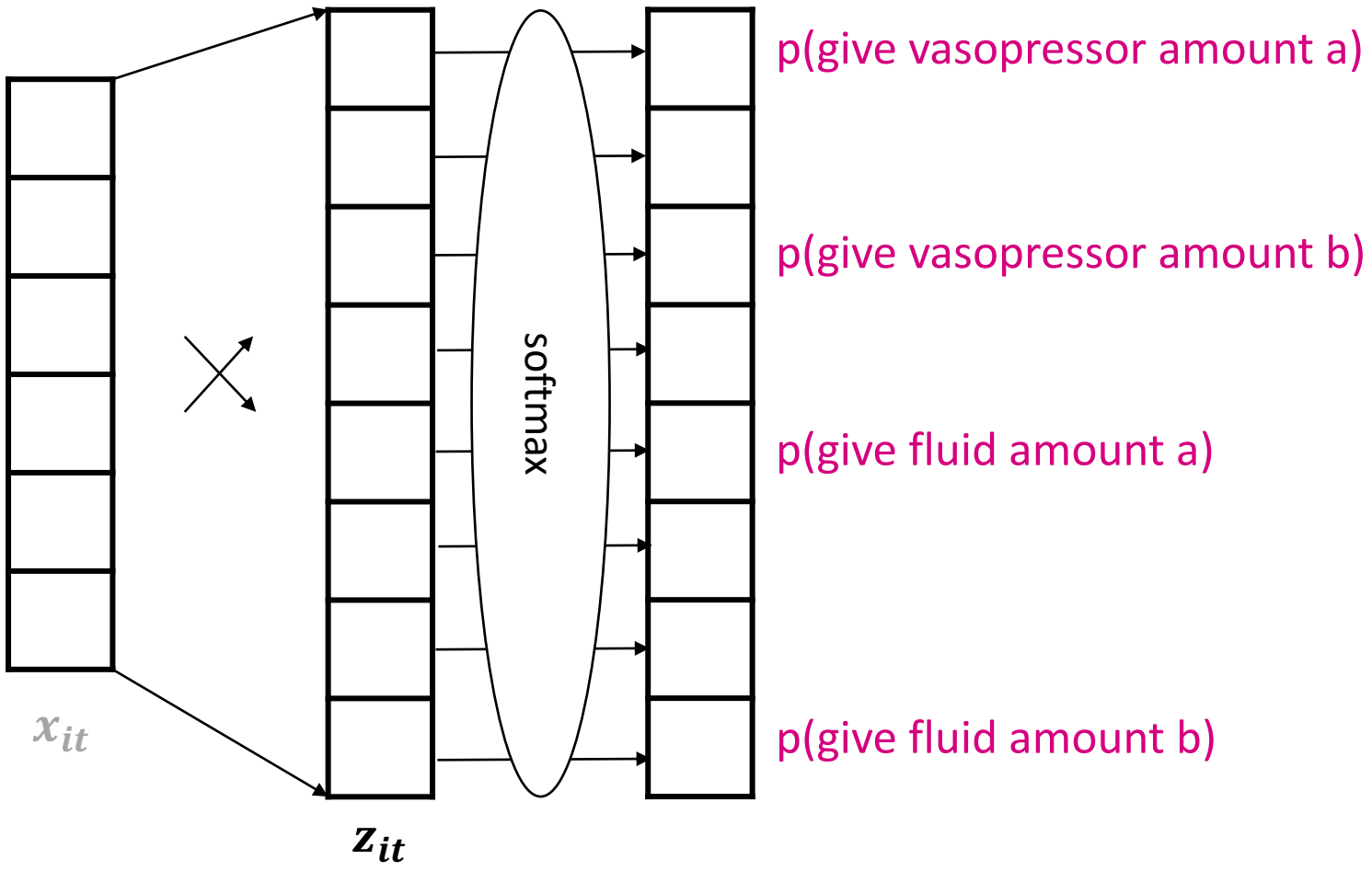


Similarities to models we've discussed

Make a series of decisions

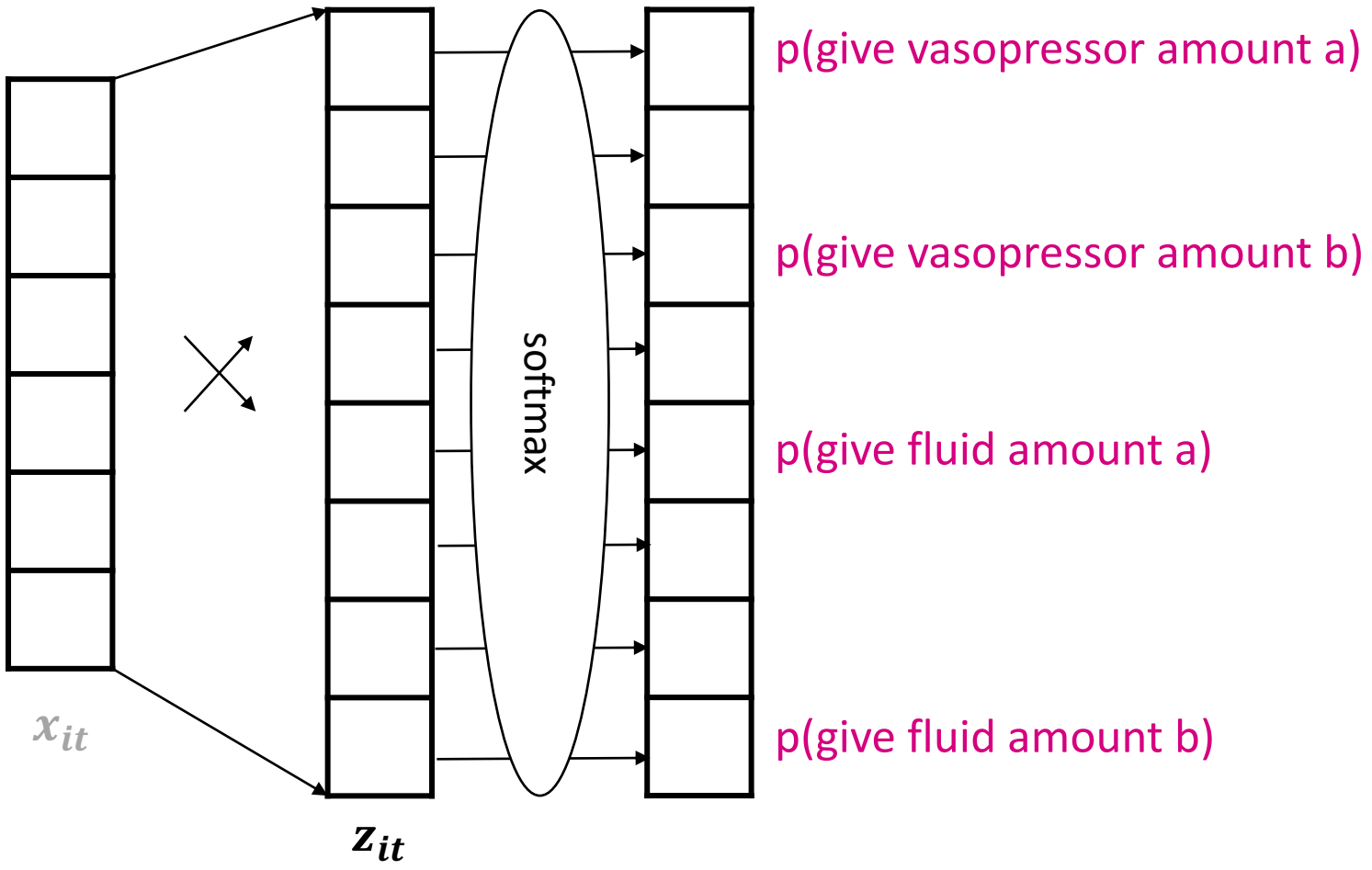
based on a set of features (state)

to maximize reward over time



Big difference: we can't directly evaluate our actions

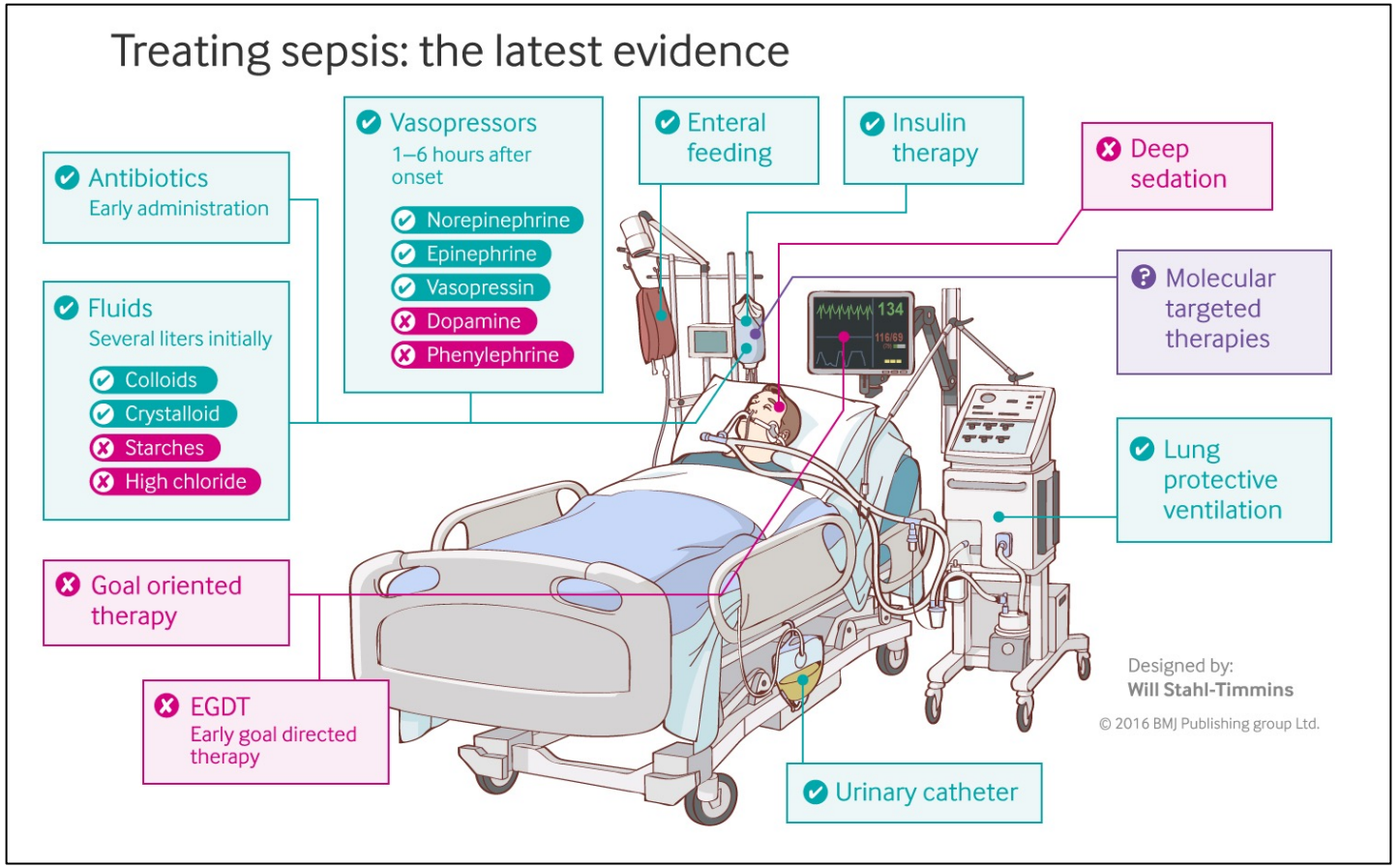
- Instead, we try things and see what works over the long-term
- Learn how our actions change the features/state, and keep track
- Try to move the system to a "good" state"



What effect do our decisions have on the patient's physiologic status?

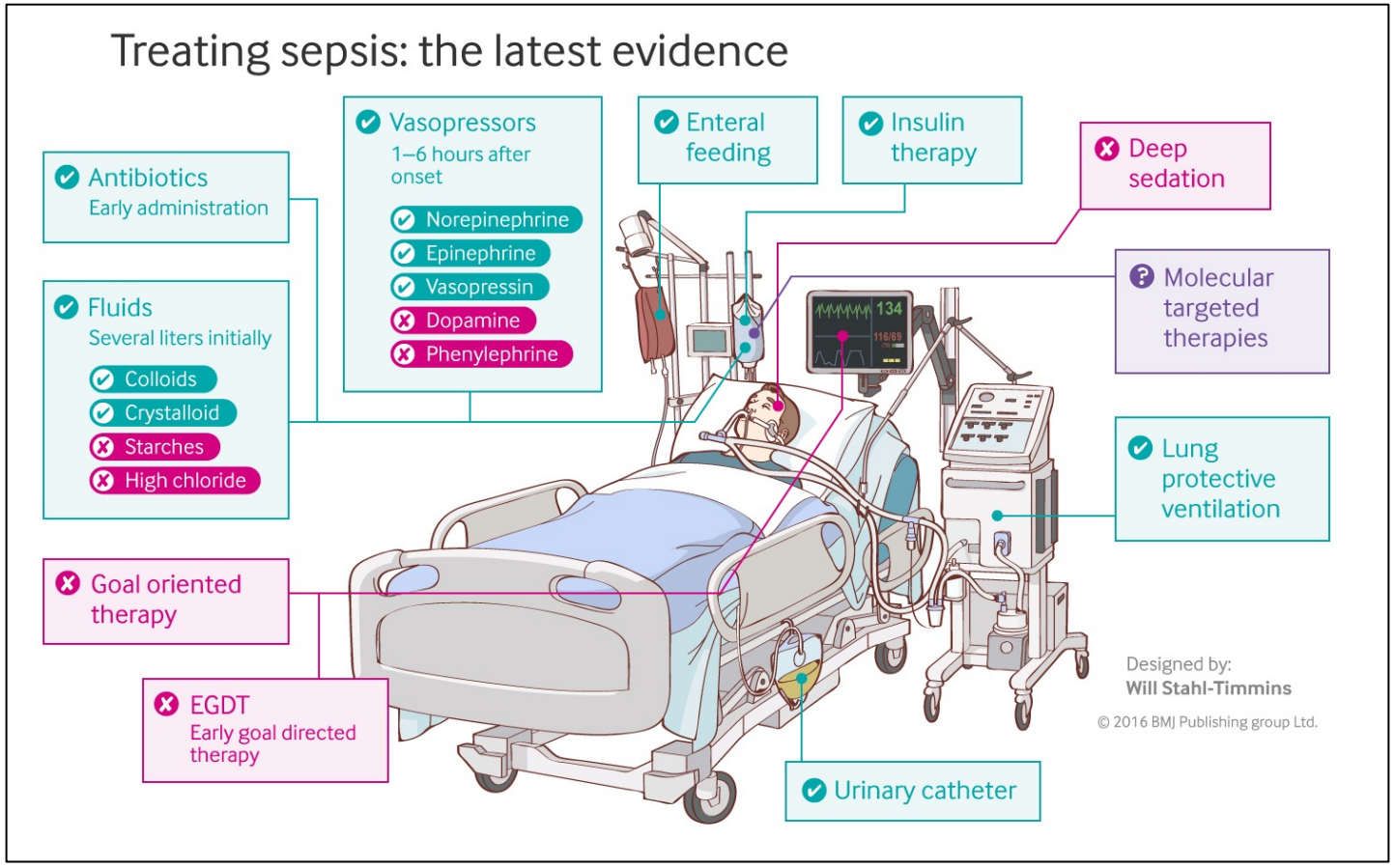
What if we give vasopressor?

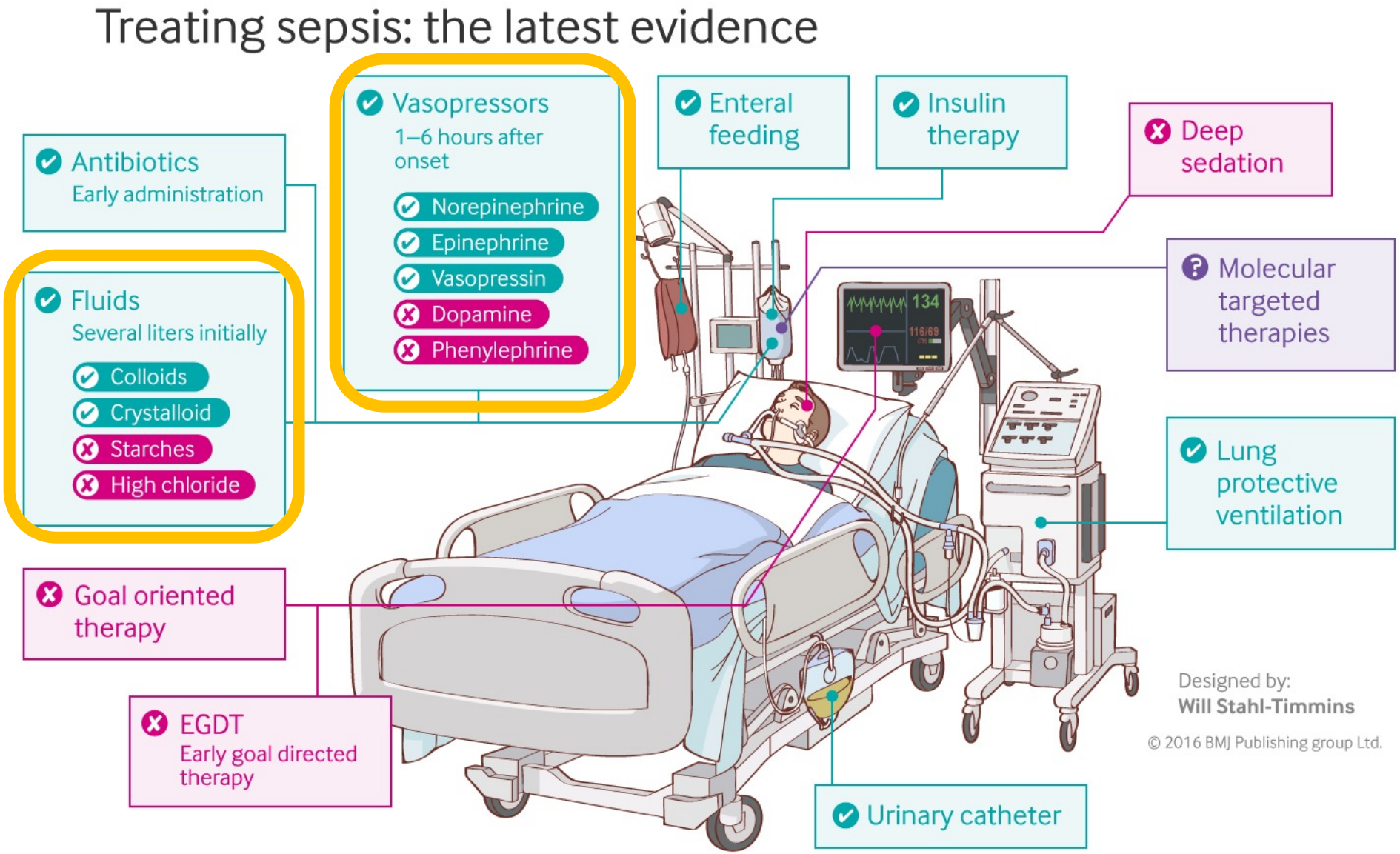
What if we give fluid?



How do changes in physiologic status, influenced by our decisions, affect the patient's mortality risk?

- Effect of actions on the features/state
- Value/reward for specific state/decision combinations
- Low lactose -> reward
- Survival -> high reward

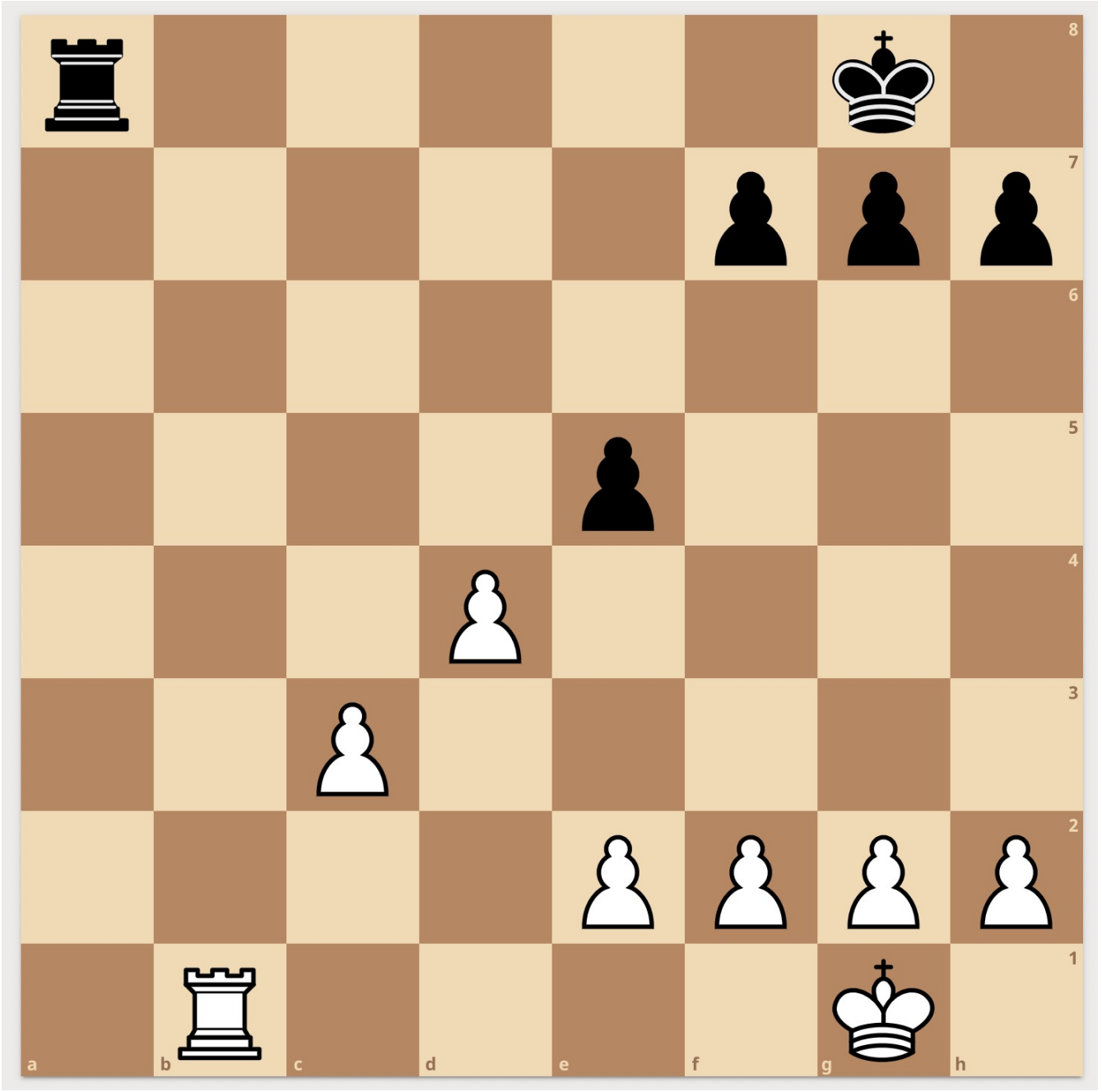




"Uncertainties still exist regarding the optimal type of fluid, the optimal volume, and the best way to monitor the response to therapy."

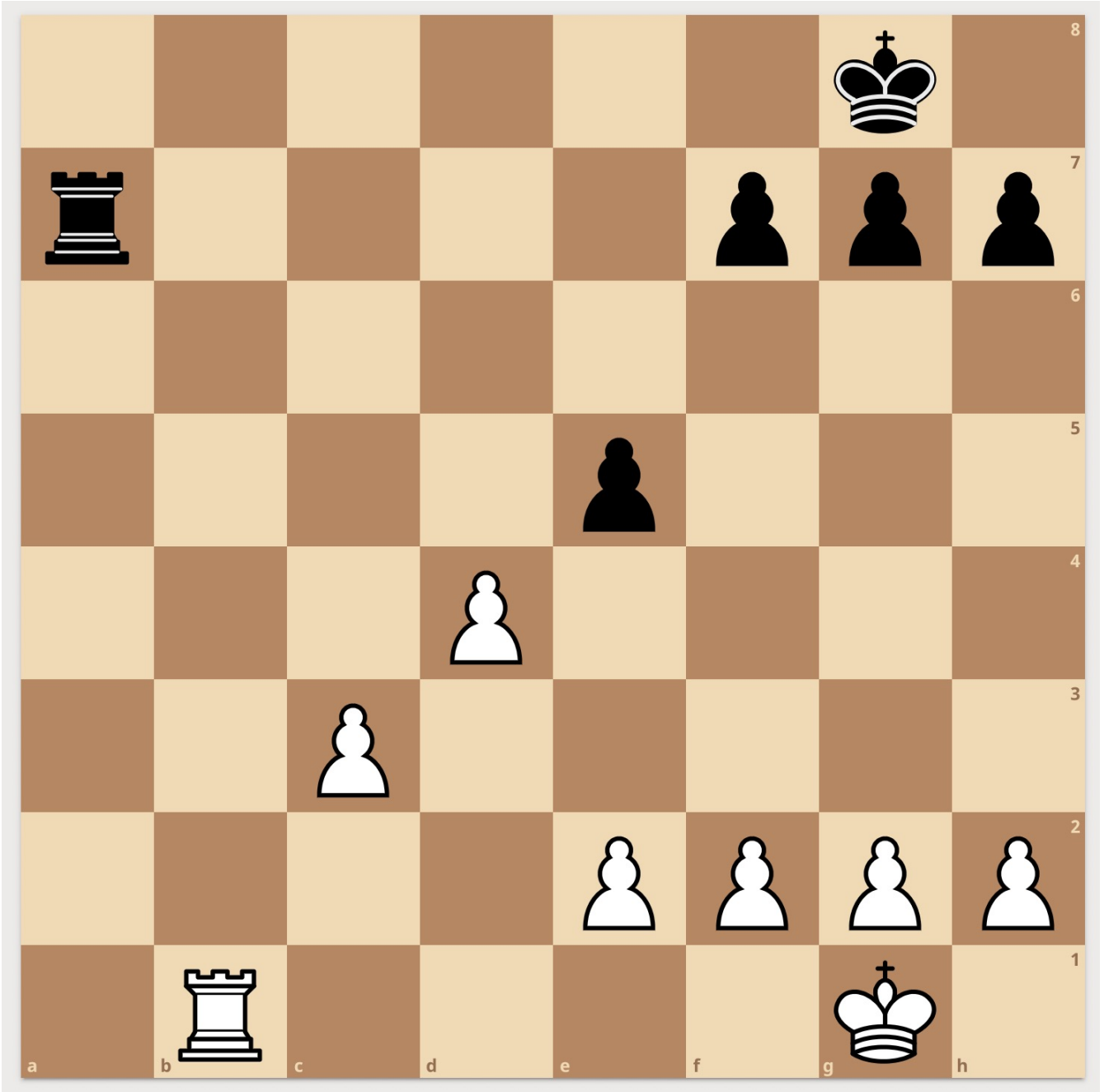
Gotts JE, Matthay MA. Sepsis: pathophysiology and clinical management. bmj. 2016 May 23;353(i1585).

Is this a safe position for black?



Is this a safe position for black?

- RL agent steers the patient toward a ‘safe’ physiologic status: one that tends not to lead to mortality



Deep Reinforcement Learning for Sepsis Treatment

Raghu A, Komorowski M, Ahmed I, Celi L, Szolovits P, Ghassemi M. arXiv:1711.09602. 2017 Nov 27

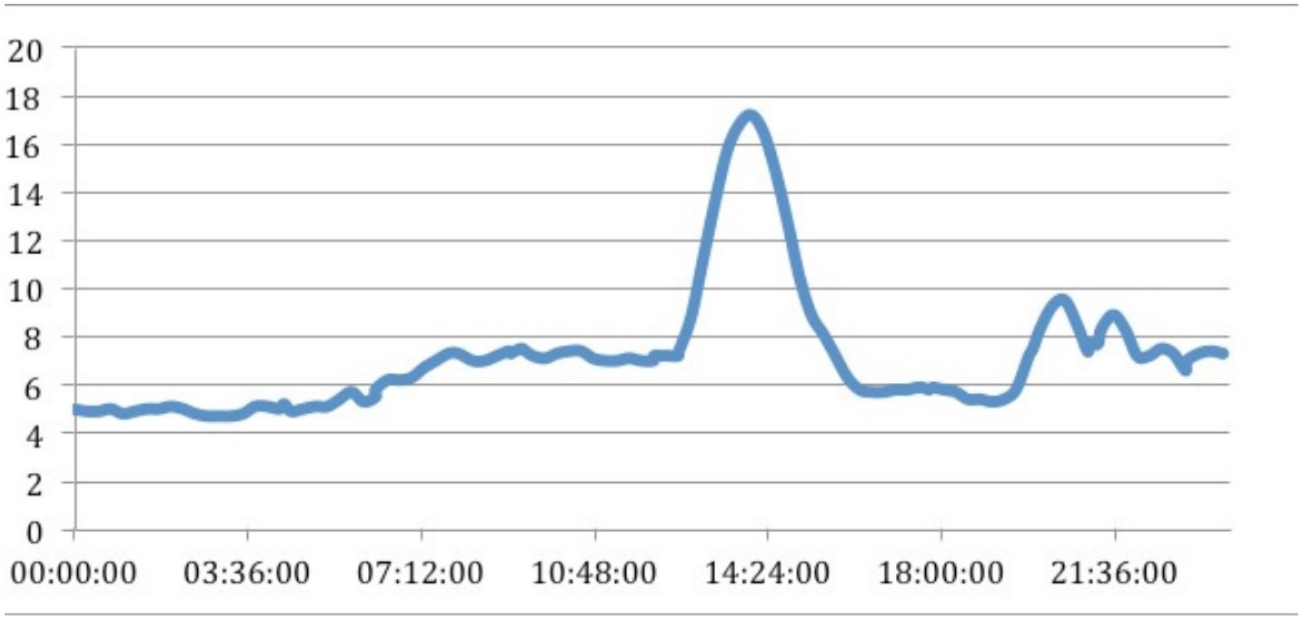
- Policy via Deep Q-Learning
- 17,898 patients from MIMIC-III

Independent validation on the Philips eICU Research Institute Database:

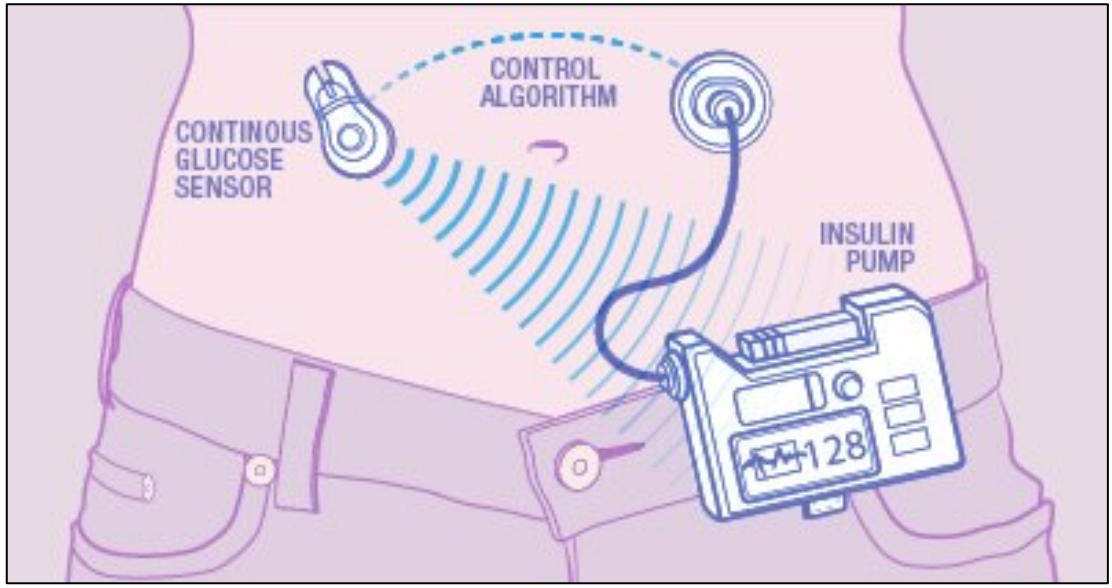
>3.3 million admissions from 2003–2016 in 459 ICUs across the US

Closed-Loop Blood Glucose Control (artificial pancreas)

Blood Glucose Readings from Continuous Glucose Monitor



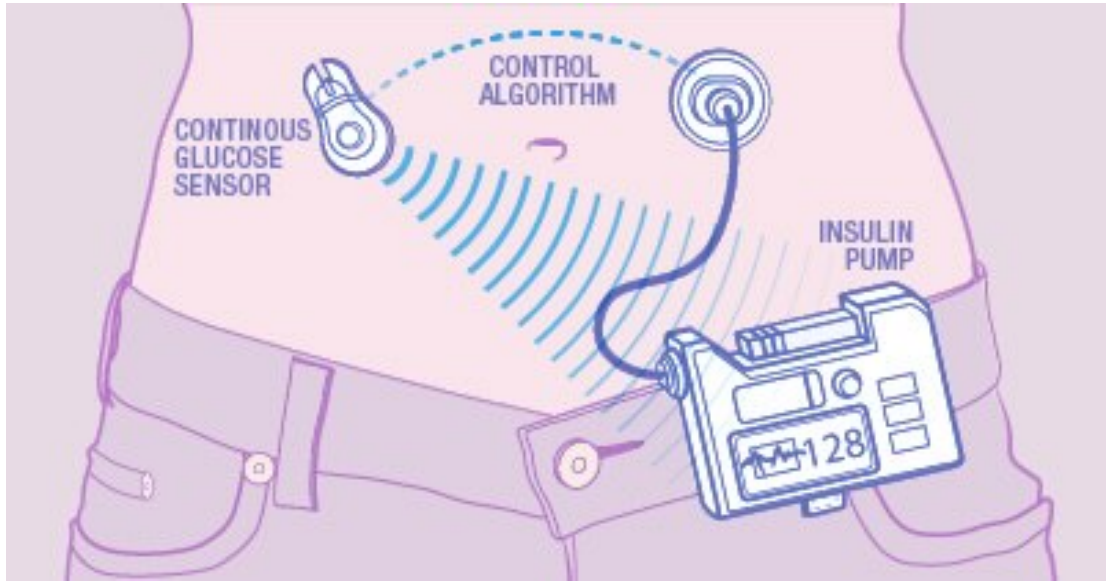
https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e



https://www.mayo.edu/research/labs/artificial-pancreas/overview

Sequential Medical Decision-Making: Artificial Pancreas

- How does an insulin bolus affect blood glucose?
- It's complicated… depends on physical activity, general health, meals, individual physiology, etc
- Goal: <u>maintain normoglycemia</u>
- Additional complication: delayed effect of actions



PLOS ONE

PUBLISH ABOUT BROWSE

OPEN ACCESS PEER-REVIEWED

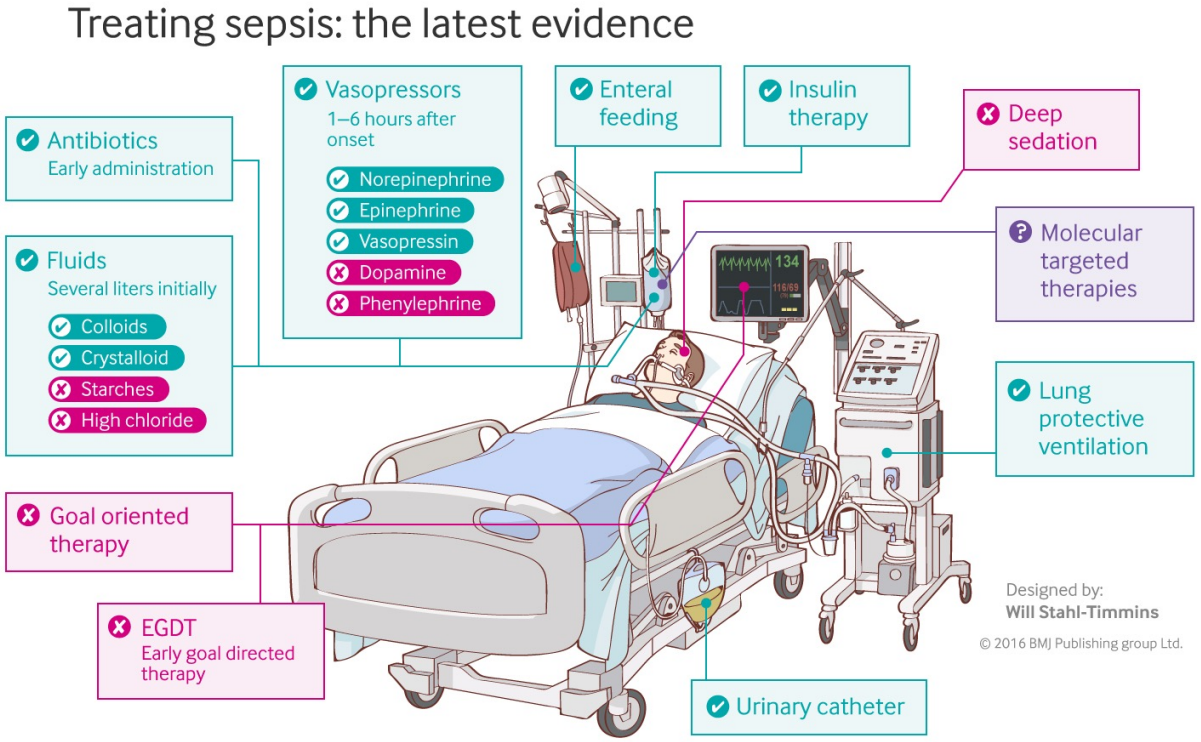
RESEARCH ARTICLE

Model-Free Machine Learning in Biomedicine: Feasibility Study in Type 1 Diabetes

Elena Daskalaki, Peter Diem, Stavroula G. Mougiakakou

Published: July 21, 2016 • https://doi.org/10.1371/journal.pone.0158722

RL's edge: keep track of everything



Clinician goals: keep the patient stable.

- central venous pressure (8-12 mm Hg)
- mean arterial pressure (65-90 mm Hg)
- urine output (0.5 mL/kg/h)
- central venous oxygen saturation (70%)

RL goals: optimize the outcome

- prevent death
- prevent organ damage

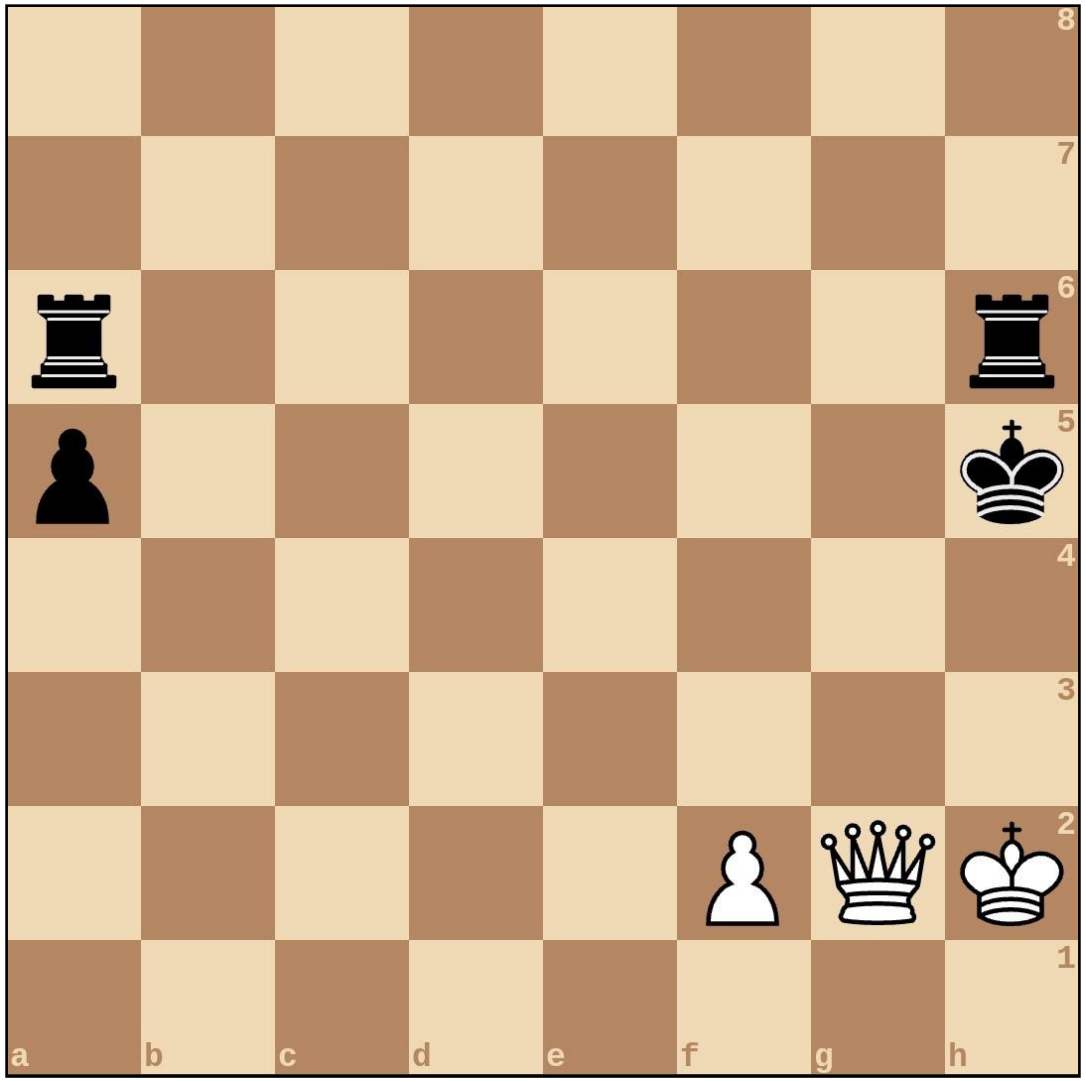
-> The RL algorithm chooses actions that maximize expected reward over time

Gotts JE, Matthay MA. Sepsis: pathophysiology and clinical management. bmj. 2016 May 23;353(i1585).

Challenge 1 for RL in Health…

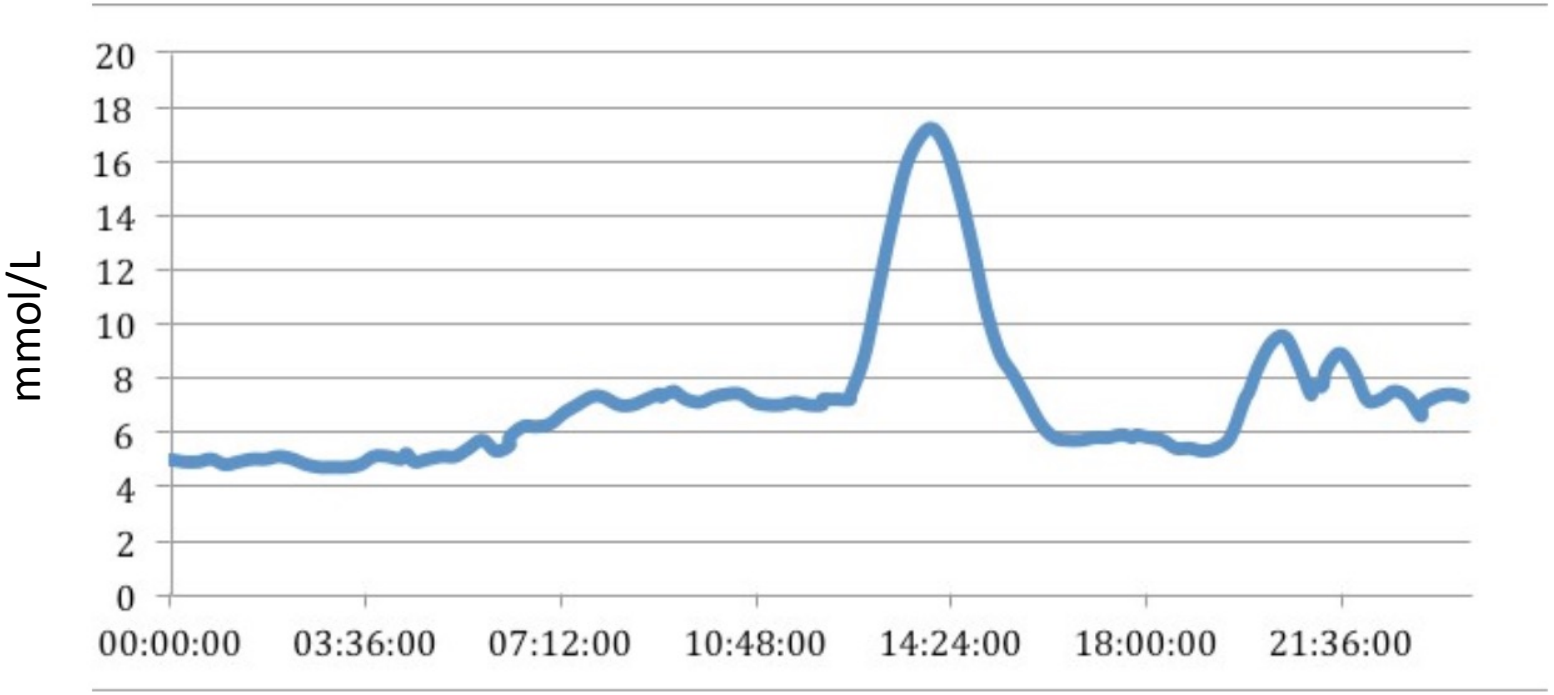
What information (features) is important?

Chess or Go: the features (state) are what you see on the board



Features x_t: Artificial Pancreas

Blood Glucose Readings from Continuous Glucose Monitor



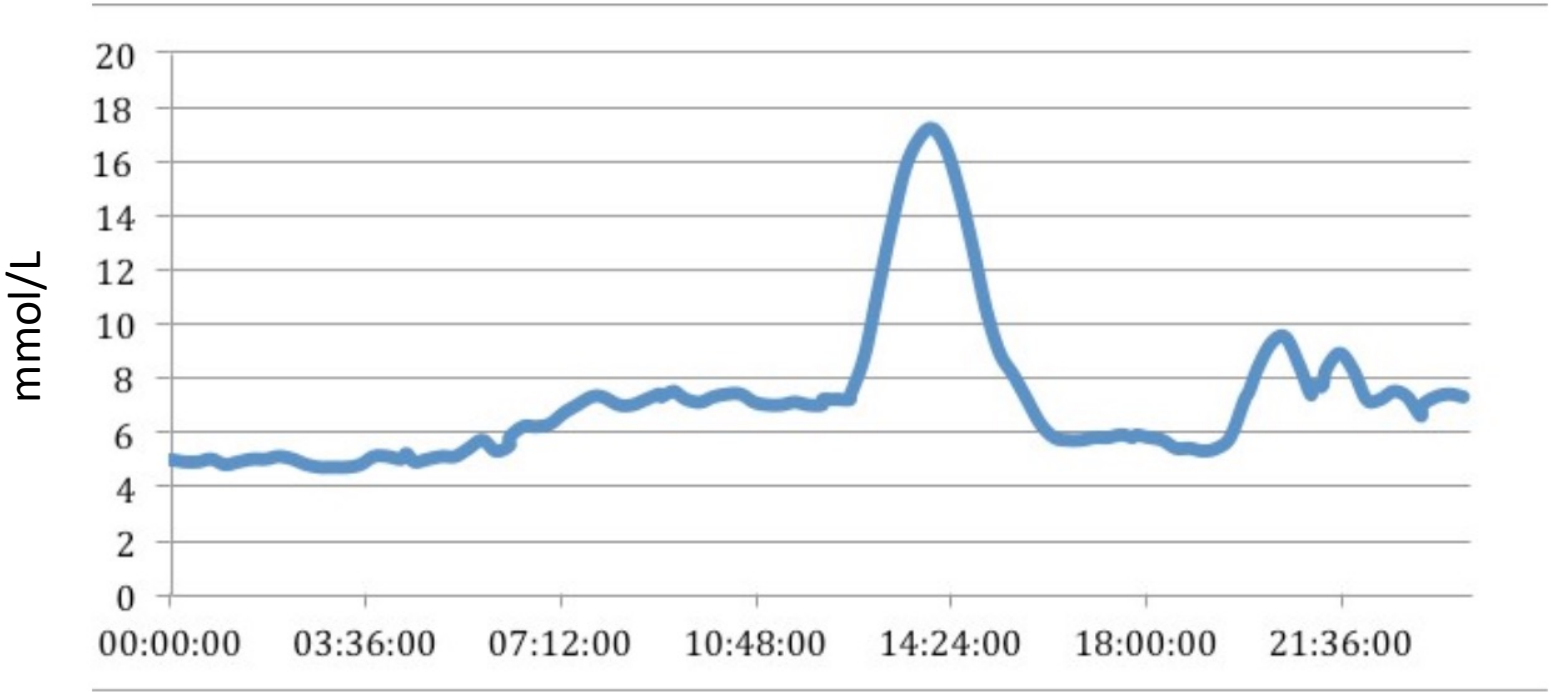
https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e

Idea 1:

The feature is the current blood glucose value

Features x_t: Artificial Pancreas

Blood Glucose Readings from Continuous Glucose Monitor



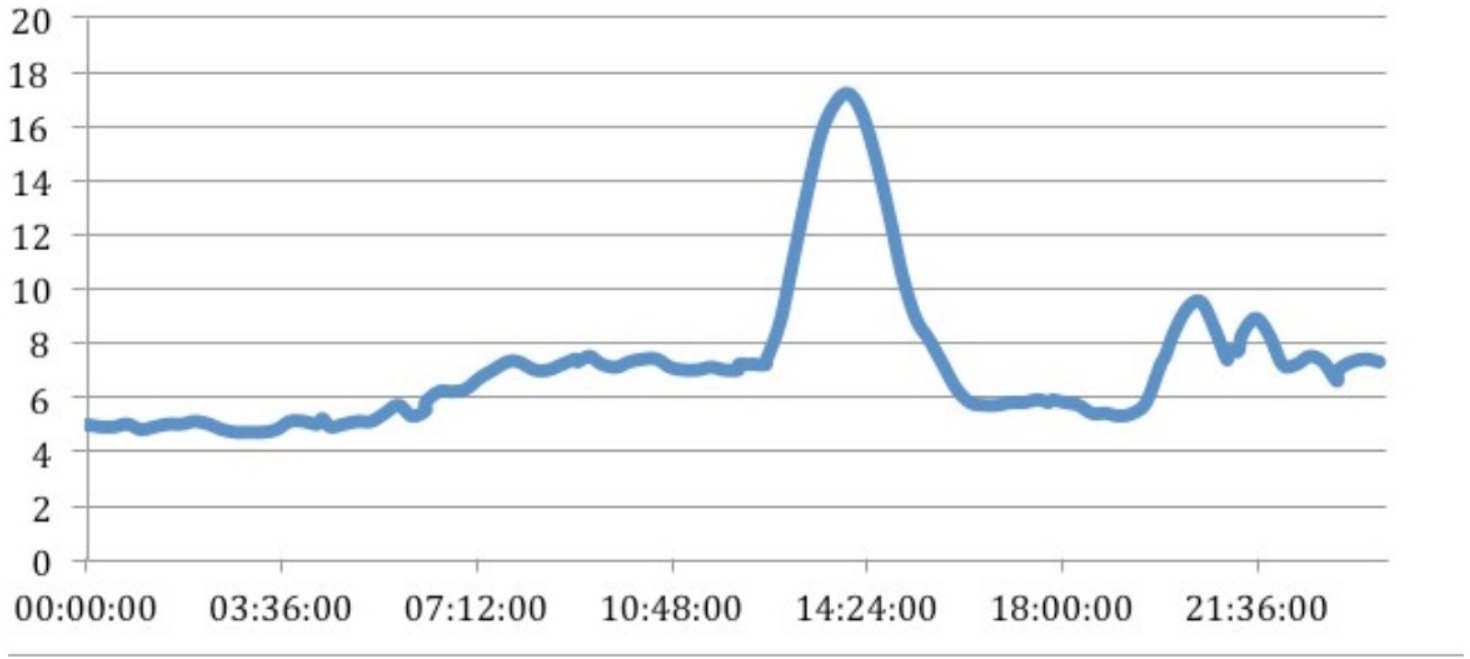
https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e

Idea 2:

The features are the current blood glucose value plus recent trends

Features x_t: Artificial Pancreas

Blood Glucose Readings from Continuous Glucose Monitor



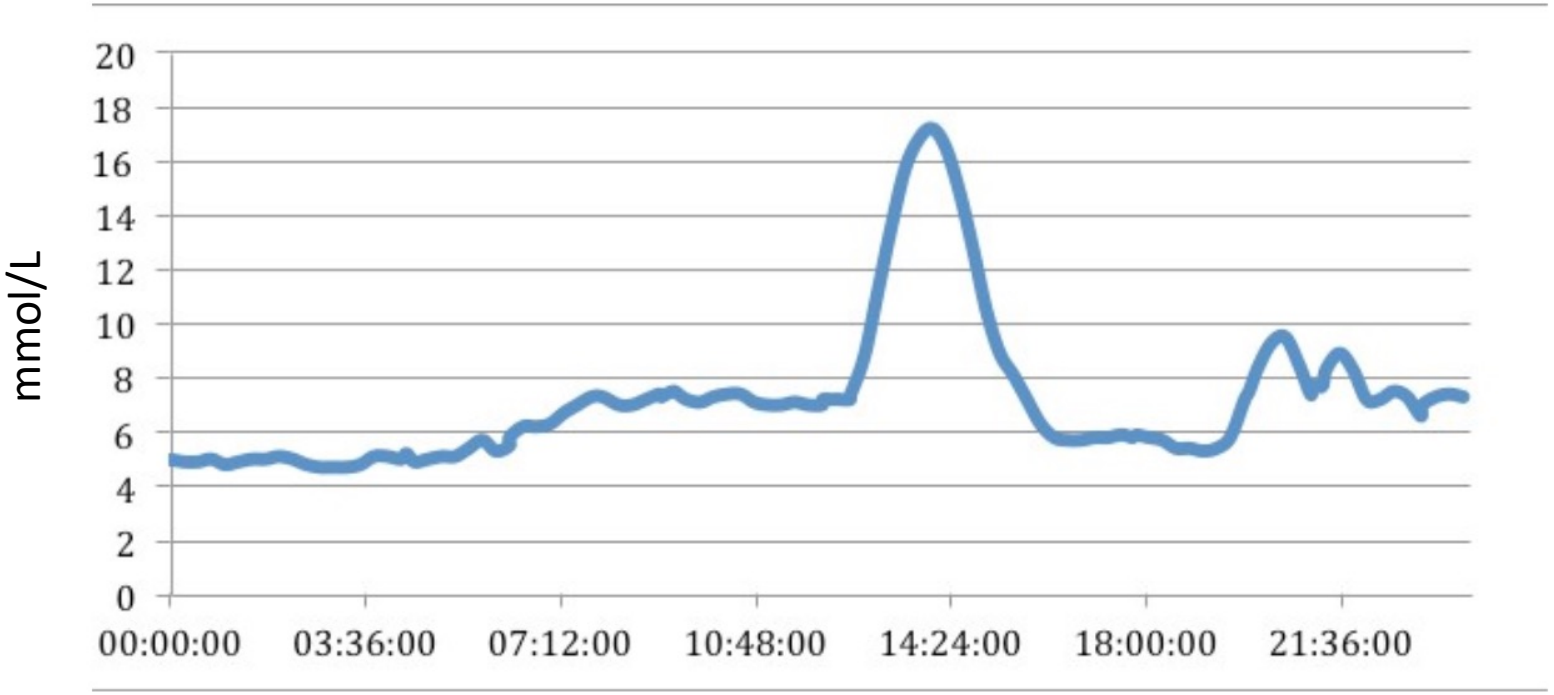
https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e

Idea 3:

The features are the current blood glucose value, recent trends, and the patient's insulin sensitivity

Features x_t: Artificial Pancreas

Blood Glucose Readings from Continuous Glucose Monitor



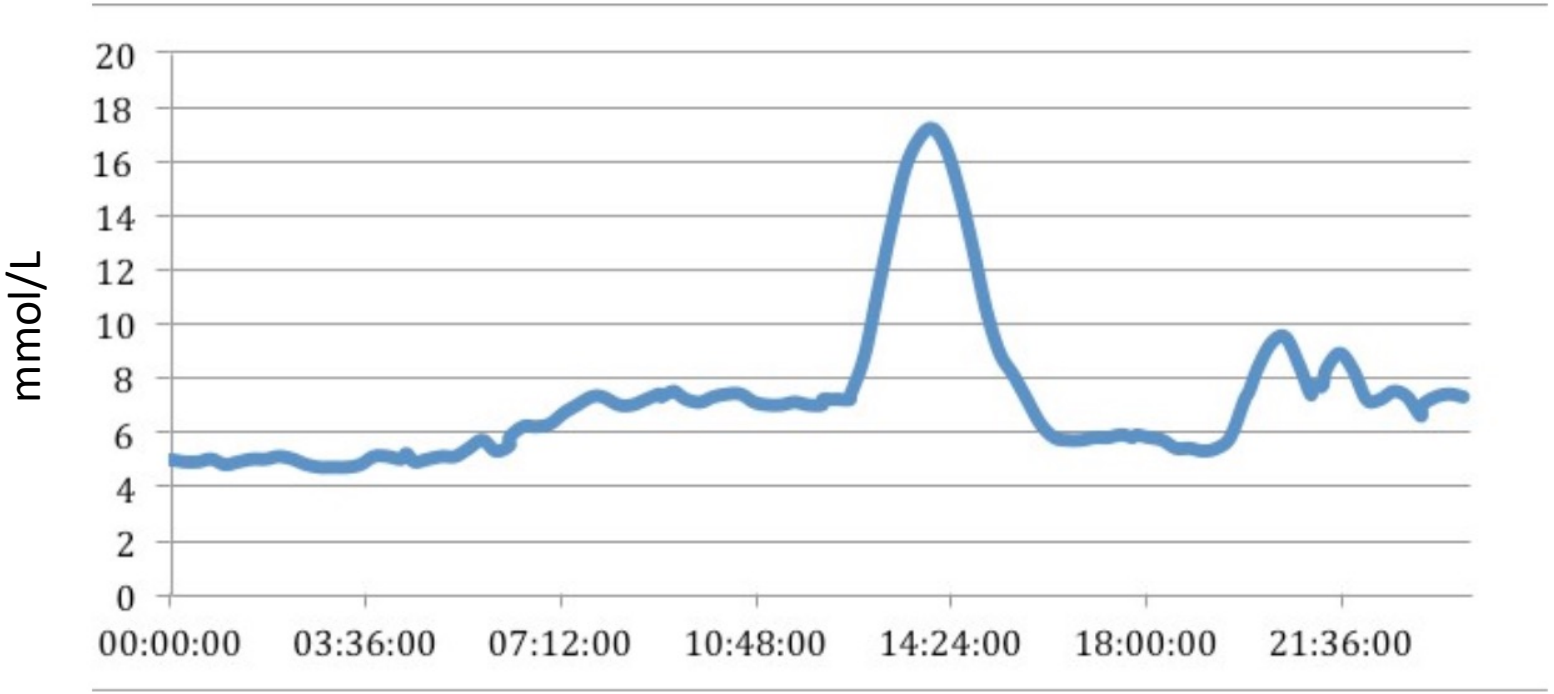
https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e

The features: all information relevant to our decision

- BG trends
- Previous insulin doses
- Patient physiology
- Recent behaviors (e.g. eating, physical activity)

Features x_t: Artificial Pancreas

Blood Glucose Readings from Continuous Glucose Monitor



https://medium.com/@justin_d_lawler/continuous-glucose-monitoring-the-first-four-weeks-7a6aa5fdb06e

Why is this critical?

Because we're making many decisions over time, each of which affects the system, errors are compounded when there is unmeasured information that affects system evolution over time

Challenge 2 for RL in Health…

We have to choose, explicitly, what the goal should be.

Challenge 3 for RL in Health…

The cost of error is high.

In RL, we typically learn “from scratch”:

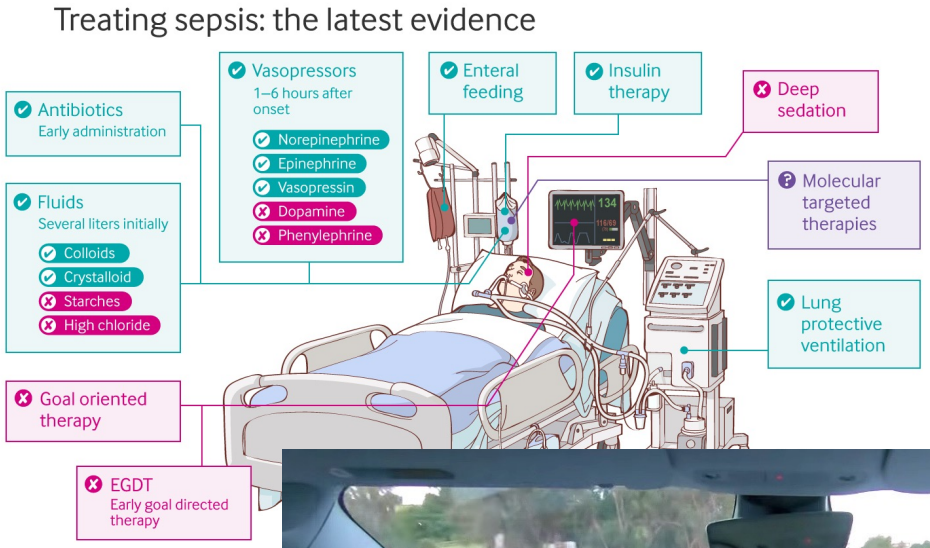
- Try things and see what works
- Initially our actions are random

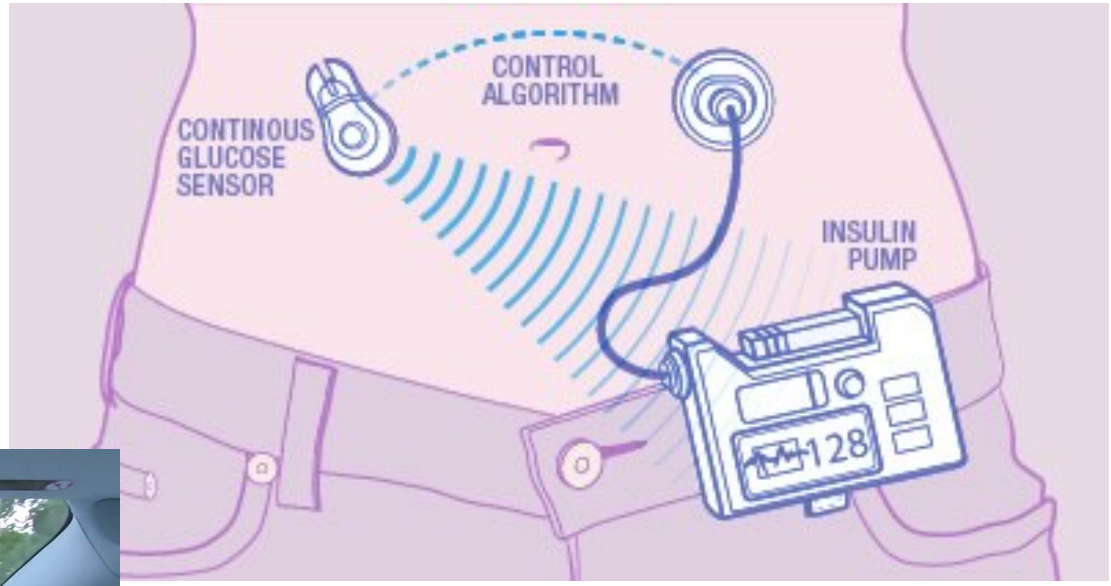
Drone Uses AI and 11,500 Crashes to Learn How to Fly

Crashing into objects has taught this drone to fly autonomously, by learning what not to do

By Evan Ackerman

Failing 11,500 times isn’t always an option





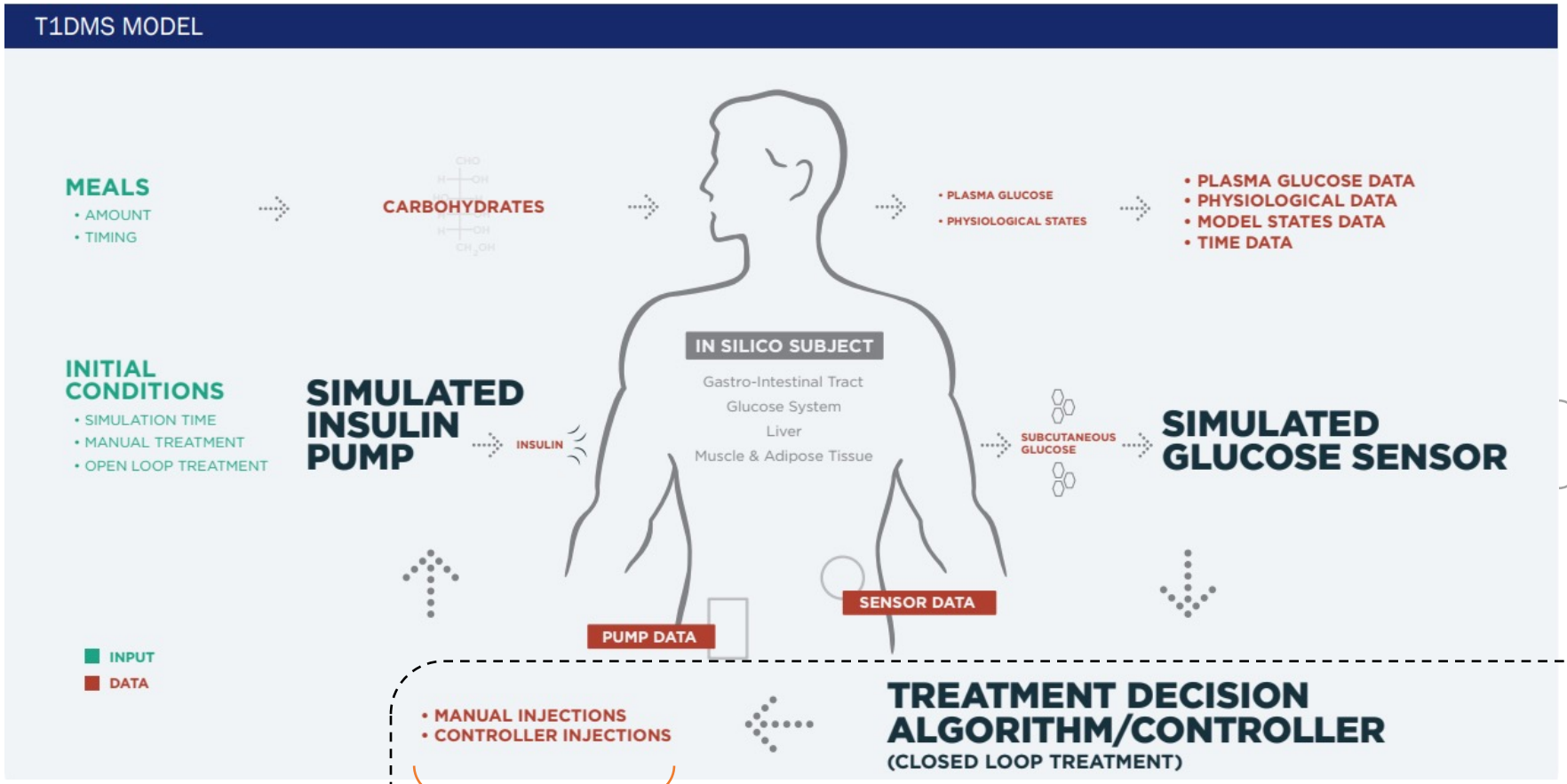
-> What are the alternatives?

A1. Learn from Observational Data

The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG and Badawi O. Scientific Data (2018).

- provides a large number of (state, action, reward) examples

A2. Learn Policy from Simulated Environment



state

actions

Reinforcement Learning Agent

A3. Learn with Expert Oversight

physiologic state

algorithm recommendation

physician approval

Open Questions for RL

1. How can we incorporate existing knowledge to avoid “starting from scratch”?

2. *Should we* avoid starting from scratch?

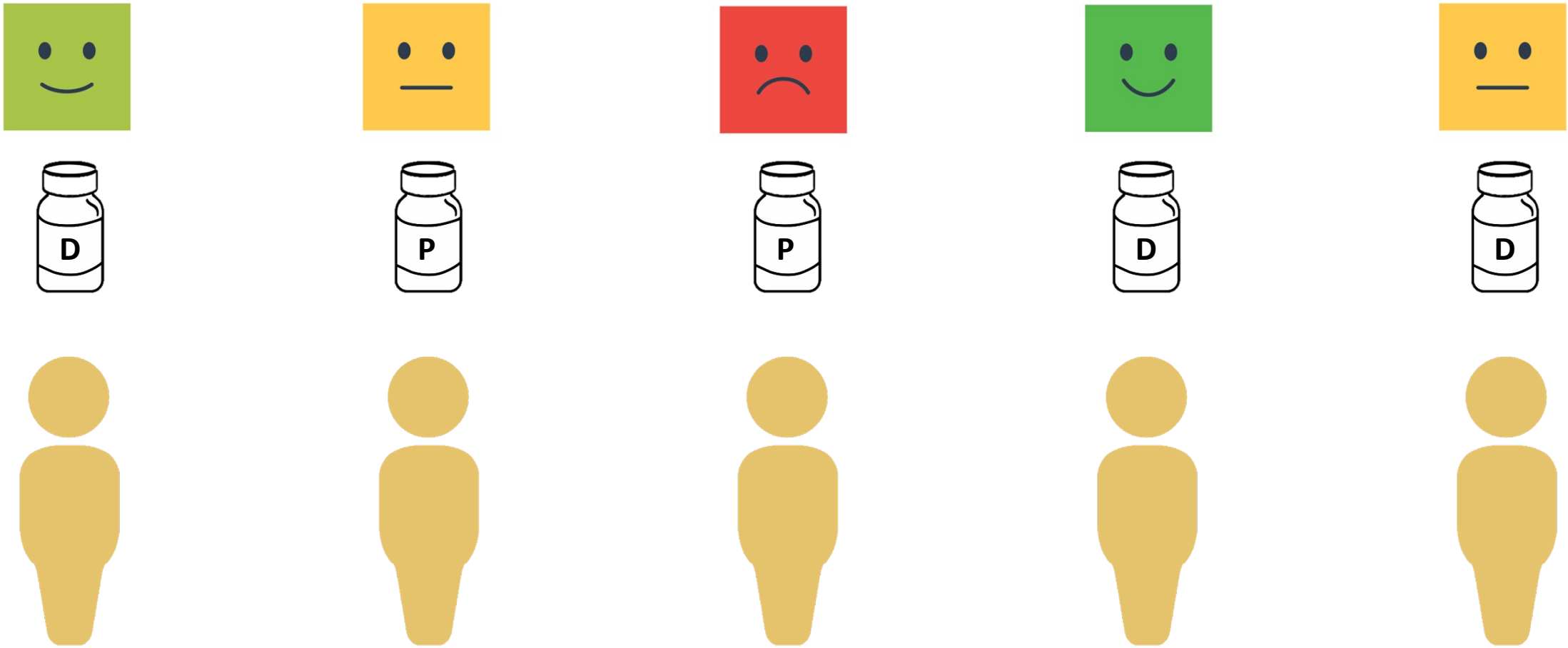
Clinical Trials and Self-Experimentation

Other Directions in RL for Medicine

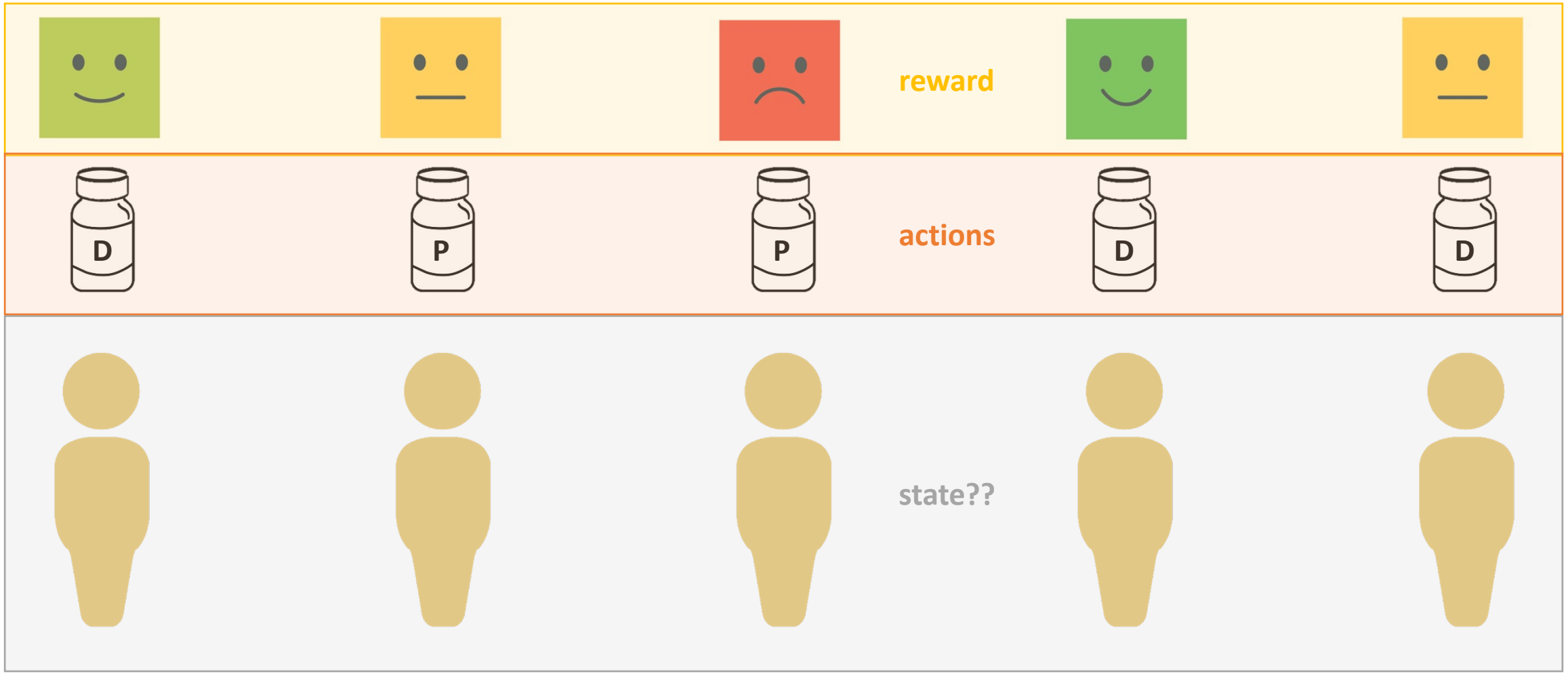
Sequential Decision-Making Problems are Everywhere in Medicine

A reinforcement learning approach to weaning of mechanical ventilation in intensive care units.
Prasad, Niranjani, et al. arXiv:1704.06300 (2017).

Suppose we are evaluating a new drug…



A special case of RL



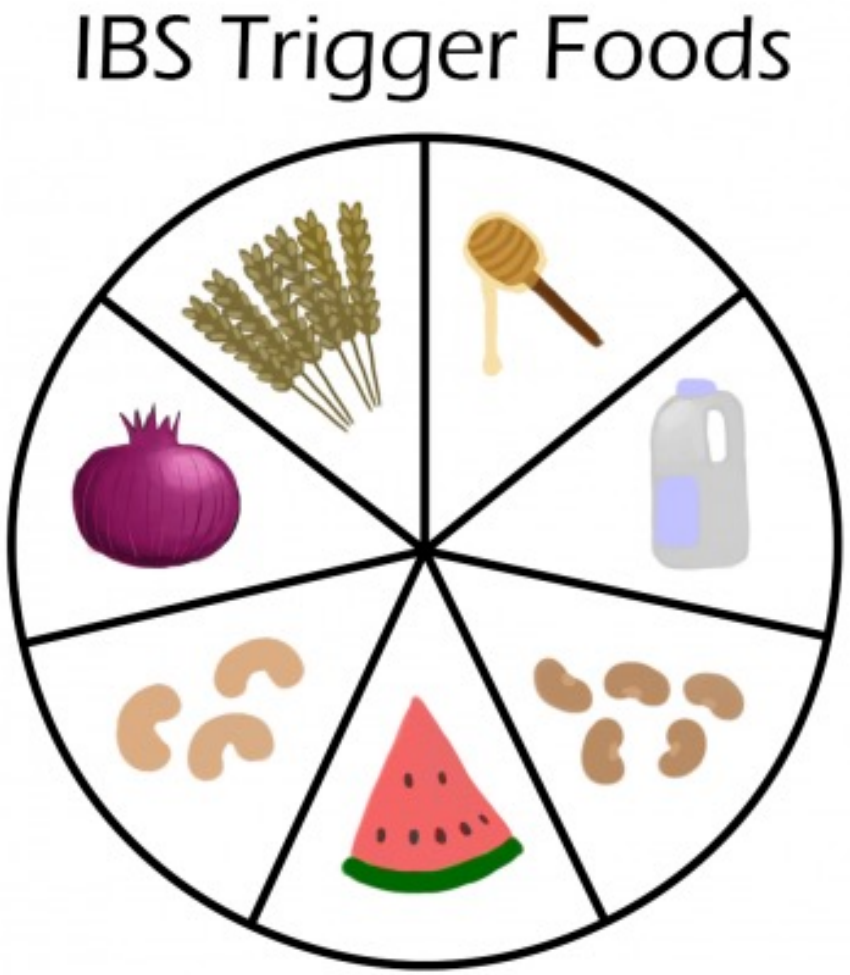
Multi-armed Bandit

Application: Optimal Allocation of Clinical Trial Participants

"An explicit assumption is the goal to treat patients effectively, in the trial as well as out. That is controversial (…)"

(Stangl, Inoue and Irony, 2012)

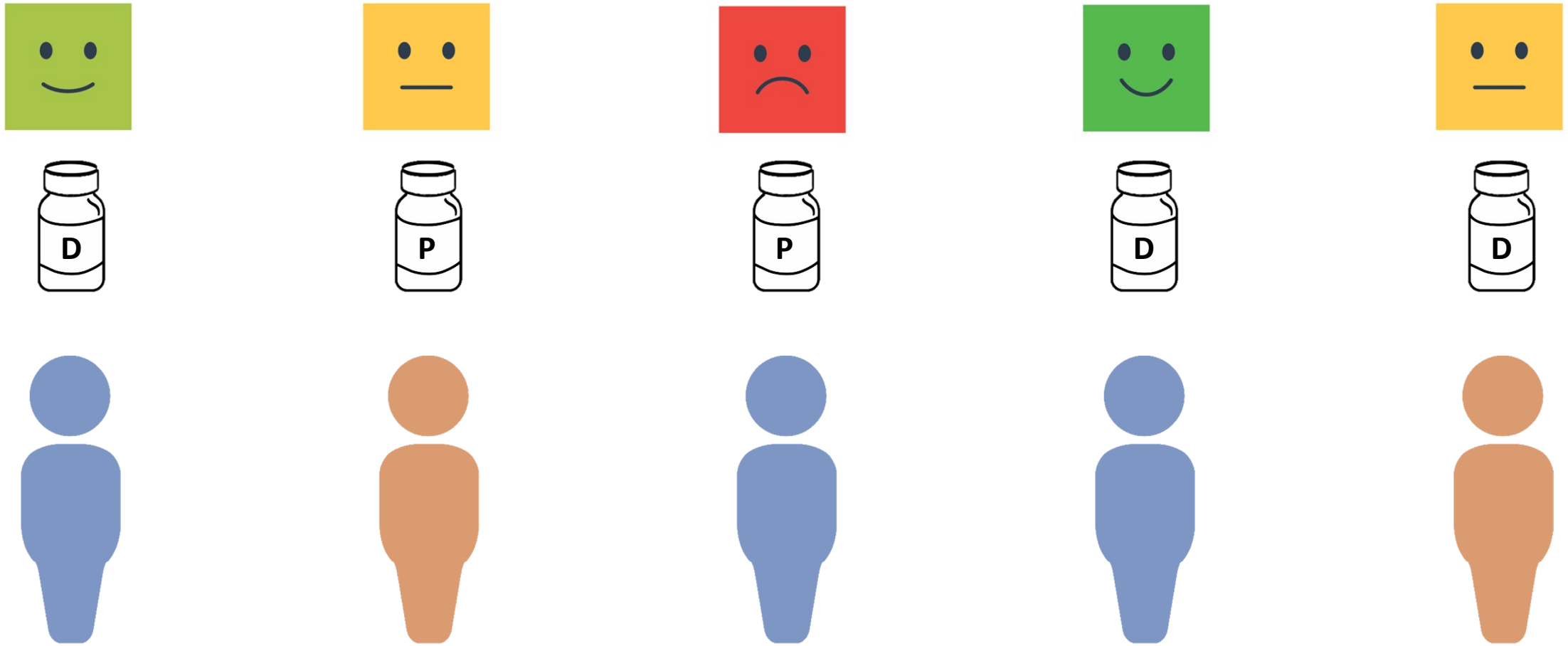
N-of-1 Trial: Identify IBS Triggers



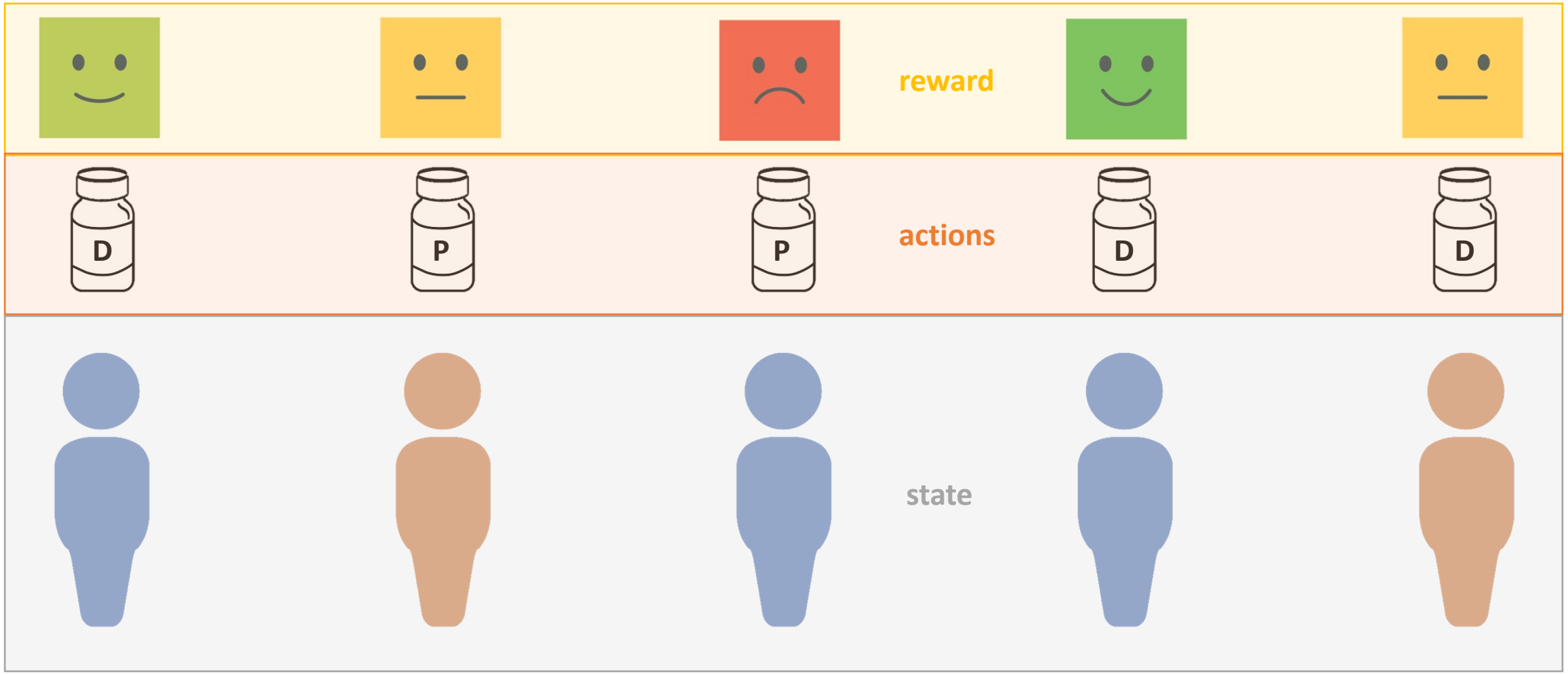
Find foods (i.e. “actions”) that minimize IBS symptoms (i.e. “reward”)

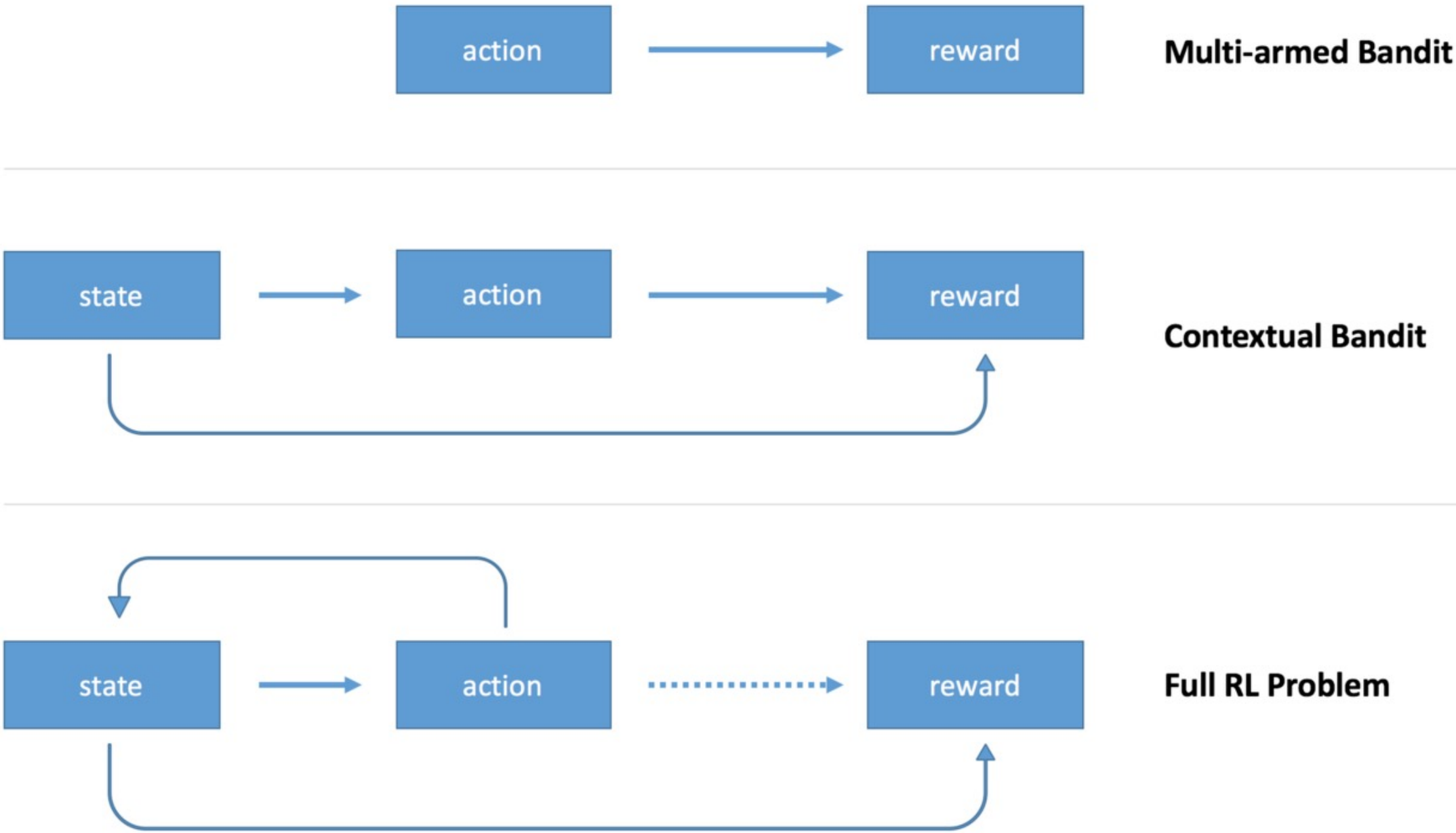
TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers.
Karkar R, Schroeder J, Epstein DA, et al.
SIGCHI Conference 2017;2017:6850-6863.

This time, we track what works for men versus women



Personalized clinical trials can be formulated as RL





From https://medium.com/emergent-future/