Supporting Preference-aware Sequential Medical Decision Making

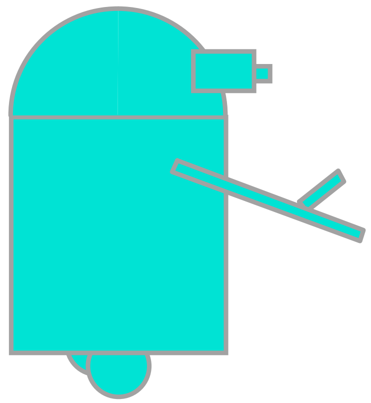
Dan Lizotte

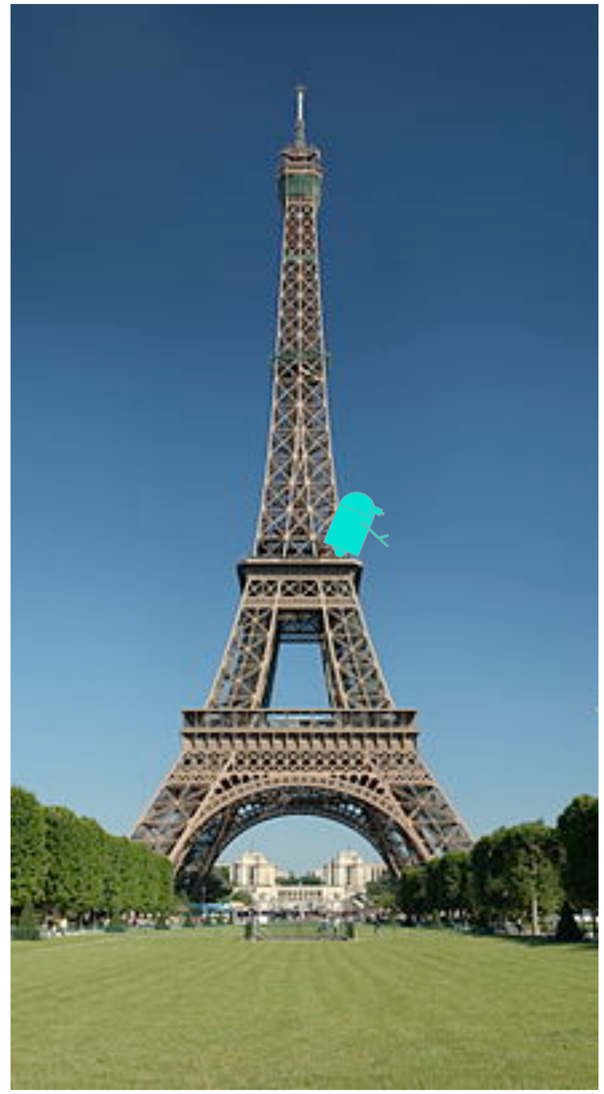
Michael Bowling, Eric Laber, Susan A. Murphy

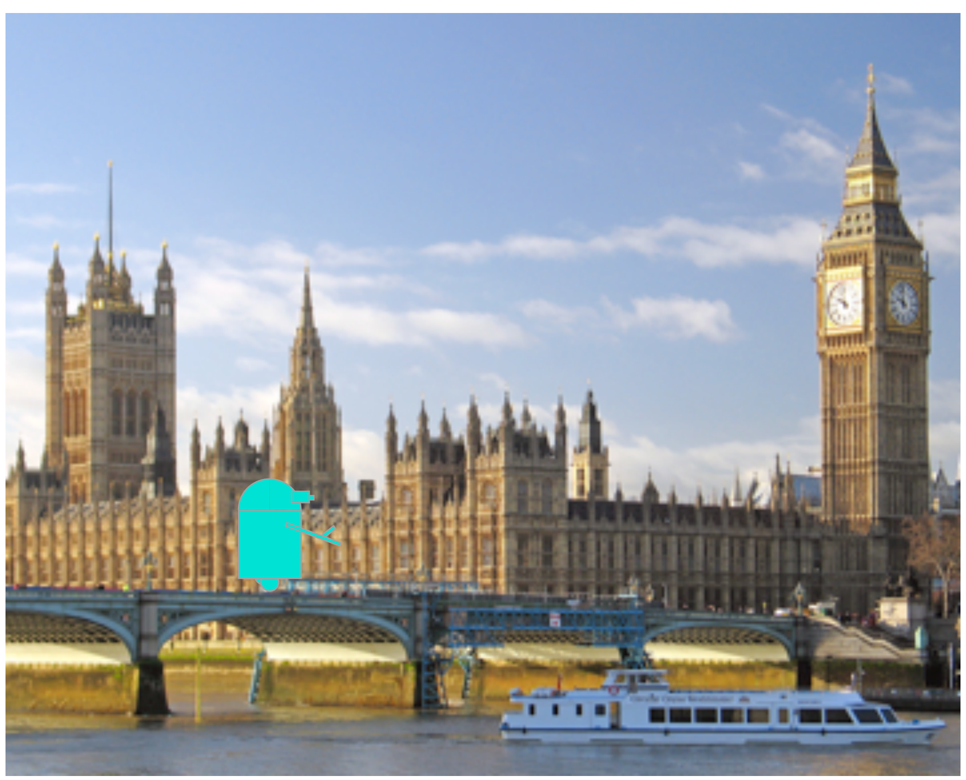
WATERLOO | CHERITON SCHOOL OF COMPUTER SCIENCE

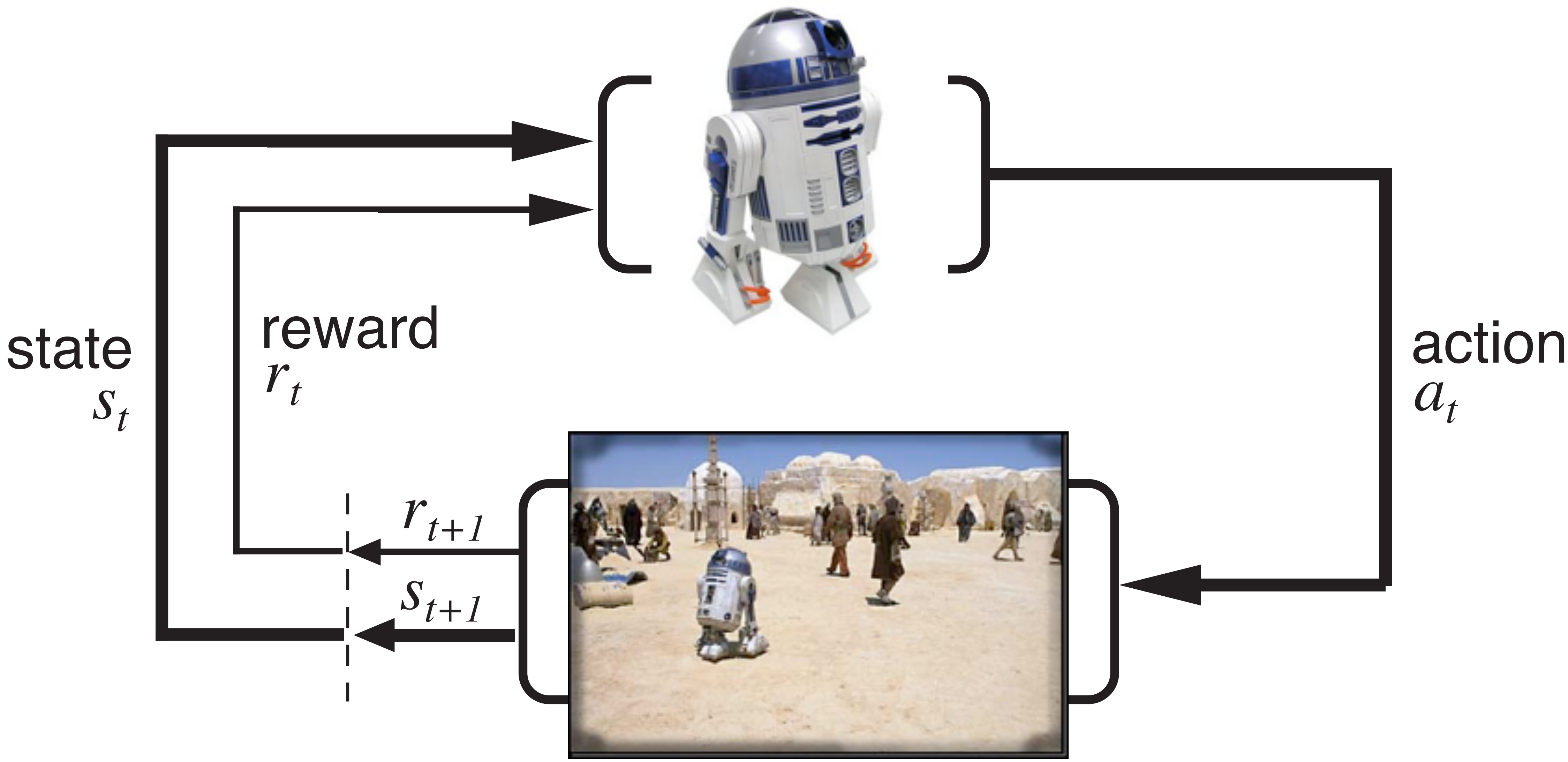
Plan

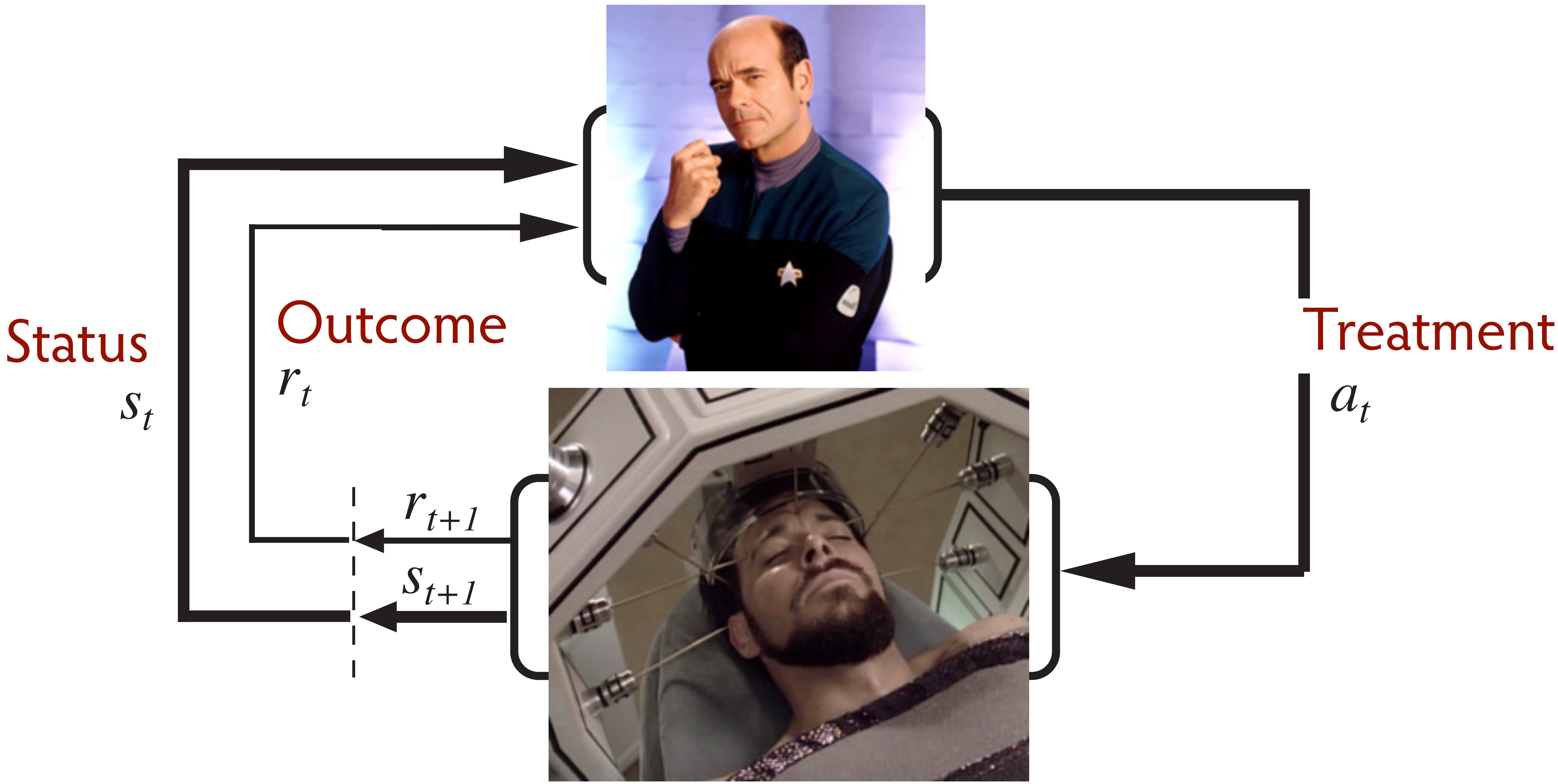
- Brief History of AI: From Autonomous Agents to Clinical Decision Support
- Argue that the Autonomous Agent approach is well-suited to but not sufficient for MUCMD
- Talk about work that tries to bring it closer



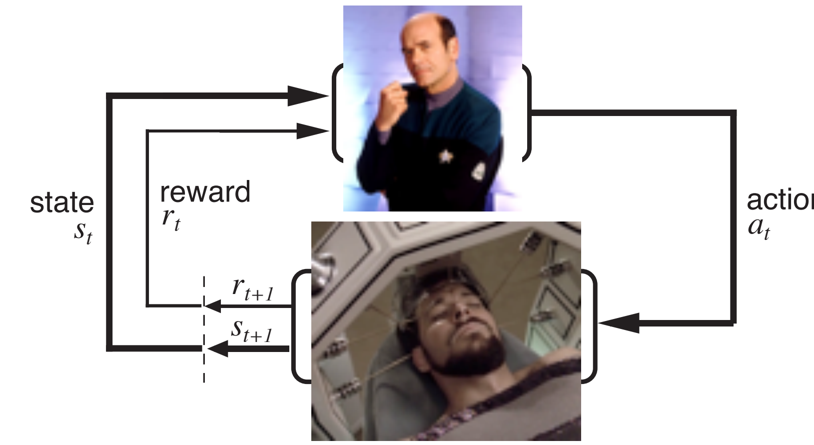





state
s_t
reward
r_t
action
a_t
r_{t+1}
s_{t+1}


Status
s_t
Outcome
r_t
Treatment
a_t
r_{t+1}
s_{t+1}

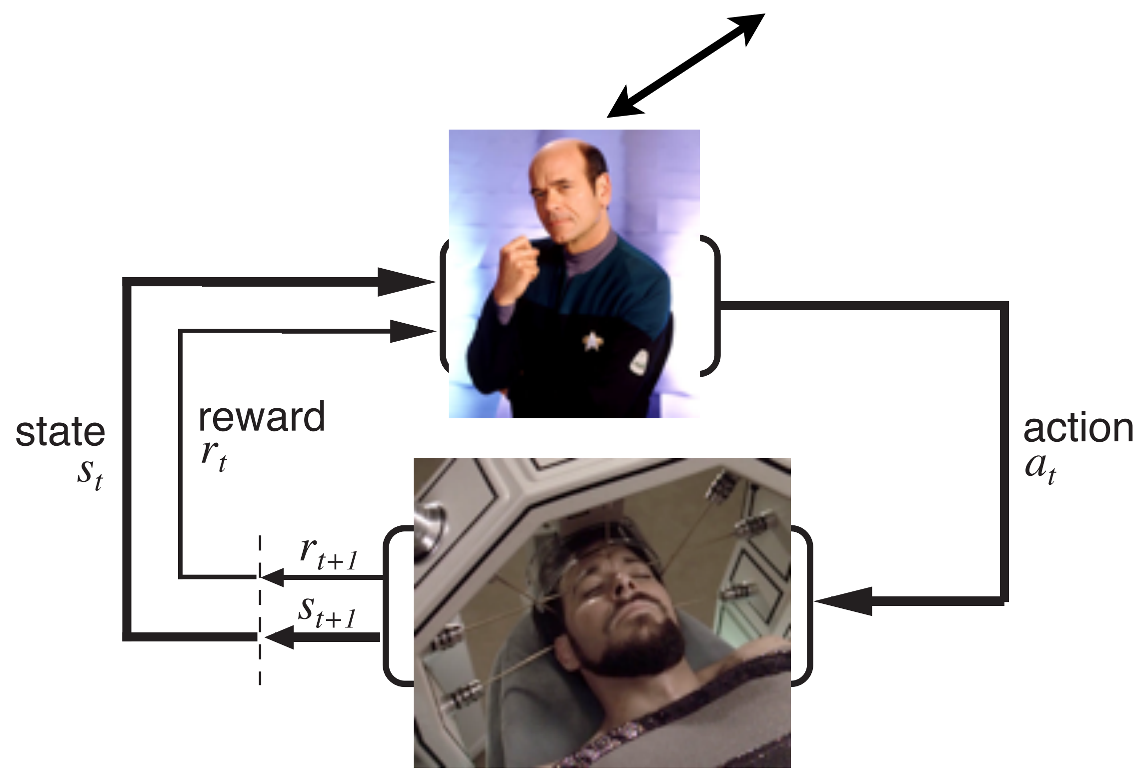
Autonomous Agent Paradigm



- Good
  - Goal is to maximize long term reward
  - Makes context-dependent decisions
  - Handles uncertain environments naturally
- Bad
  - Doesn't give rigorous confidence measures*
  - Assumes complete state information
    (or that you know what you don't know)
  - Relies relies on "correct" reward specification

Decision Support Agent

- Assumes complete state information
  (or that you know what you don't know)



- Decision Support Agent **still** relies on "correct" reward specification

The “Reward Hypothesis”

- “That all of what we mean by goals and purposes can be well thought of as maximization of the expected value of the cumulative sum of a received scalar signal (reward).” -- Rich Sutton

Competing Outcomes

- Different antipsychotics have different effects on symptom reduction and weight gain
- They also have different effects on different individuals
- What should we optimize?

From decision making to decision support

- Relies relies on "correct" reward specification
- How can we mitigate this?
  - Preference Elicitation (sort of)
  - Preference Revealing (DL,Bowling,Murphy)
  - Multi-outcome Screening (DL,Ferguson,Laber)

Background: Treatment Policies

- Treatment policies attempt to operationalize sequential clinical decision making
- Sequence of decision rules, one for each decision point.
  - Input: patient information
  - Output: a recommended treatment.
- One goal: find the treatment policy that maximizes the expectation of a chosen clinical outcome.

Formalism

At each decision point from t = 1 to t = T, a state is observed, an action is taken, and subsequently, a reward is observed.

State: s_t — Current knowledge about the patient needed for decision making. May include past treatments and observations.

Action: a_t — Treatment action. The set of available actions may change over time.

Reward: r_t — A **scalar outcome** based on observation of the patient's response to treatment, coded so that higher values are preferred.

Q-Learning

Use **regression**: $Q(s_T, a_T) \approx E[R_T \mid s_T, a_T]$

Recommended action for state s_T is $\text{argmax}_a\, Q(s_T, a)$

Value of a state is given by $V(s_T) = \max_a Q(s_T, a)$

For T-1, maximize expectation of current reward plus future reward assuming we act optimally.

$Q(s_{T\text{-}1}, a_{T\text{-}1}) \approx E[R_{T\text{-}1} + V(S_T) \mid s_{T\text{-}1}, a_{T\text{-}1}]$

Recommended action for $s_{T\text{-}1}$ is $\text{argmax}_a\, Q(s_{T\text{-}1}, a)$...

Preference Elicitation

Suppose D different rewards are important for decision-making,

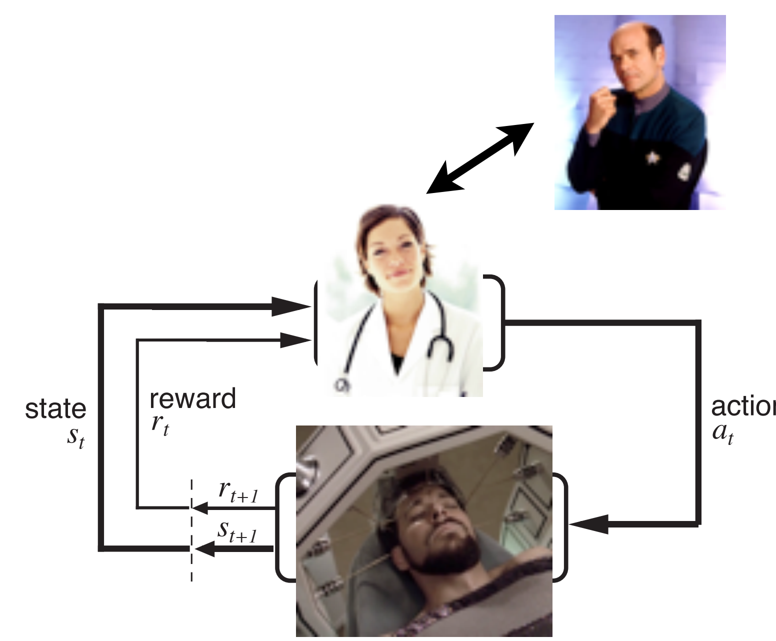
$$r_{[1]}, r_{[2]}, \ldots, r_{[D]}$$

Assume each person has a function f that takes these and gives utility, that person's happiness given any configuration of the $r_{[i]}$ expressed as a **scalar** value. We could use this as our new reward!

Preference Elicitation attempts to figure out an **individual's** f.

Preference Elicitation

1. Determine preferences of the decision-maker
2. Construct reward function from "basis rewards" (different outcomes)
3. Compute the recommended treatment, e.g. with Q-learning



Preference Elicitation

One way: Assume f has a nice form:

$$f(r_{[1]}, r_{[2]}, \ldots, r_{[D]}) = \delta_{[1]}r_{[1]} + \delta_{[2]}r_{[2]} + \ldots + \delta_{[D]}r_{[D]}$$

Then Preference Elicitation figures out the δ, or weights, an individual attaches to different rewards. How?

The values $\delta_{[i]}$ and $\delta_{[j]}$ defines an exchange rate between $r_{[i]}$ and $r_{[j]}$.

Preference Elicitation

$$f(r_{[1]}, r_{[2]}, \ldots, r_{[D]}) = \delta_{[1]} r_{[1]} + \delta_{[2]} r_{[2]} + \ldots + \delta_{[D]} r_{[D]}$$

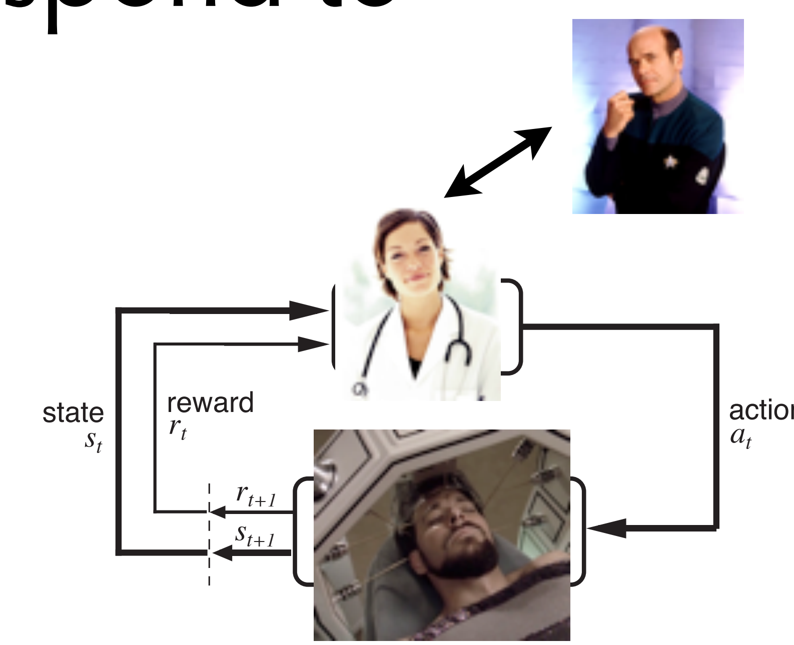
"If I lost $\delta_{[j]}$ units of $r_{[i]}$,
but I gained $\delta_{[i]}$ units of $r_{[j]}$,
I would be equally happy."

Preference elicitation asks questions like:
"If I took away 5 units of $r_{[i]}$, how many units of $r_{[j]}$ would you want?"

Once f is known, standard single-outcome methods can be applied.

Preference Elicitation

- Are the questions based in reality?
- Even if they are, can the decision-maker answer them?
- How will the decision-maker respond to "I know what you want." ?



Preference Revealing

1. ~~Determine the preferences of the decision-maker~~
2. Compute the recommended treatment **for all possible preferences (**δ**)**
3. **Show**, for each action, what preferences are consistent with that action being recommended

Preference Revealing

Benefits

No reliance on preference **elicitation**

Facilitates **deliberation** rather than imposing a single recommended treatment

Information still **individualized** through patient state

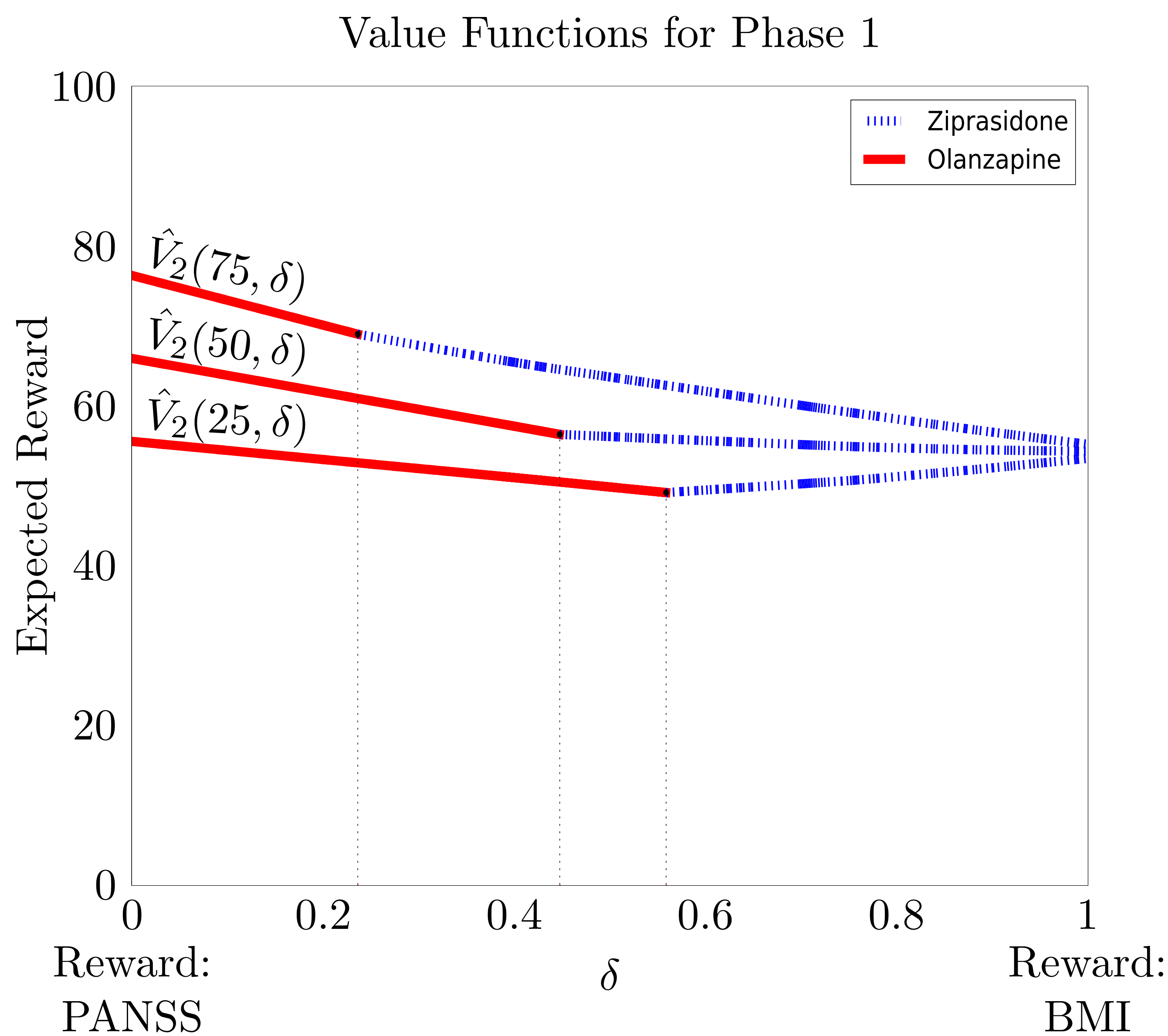
Treatments that are not suggested for any preference are implicitly **screened**

Positive And Negative Syndrome Scale

vs.

Body Mass Index

Phase 1



Positive And Negative Syndrome Scale

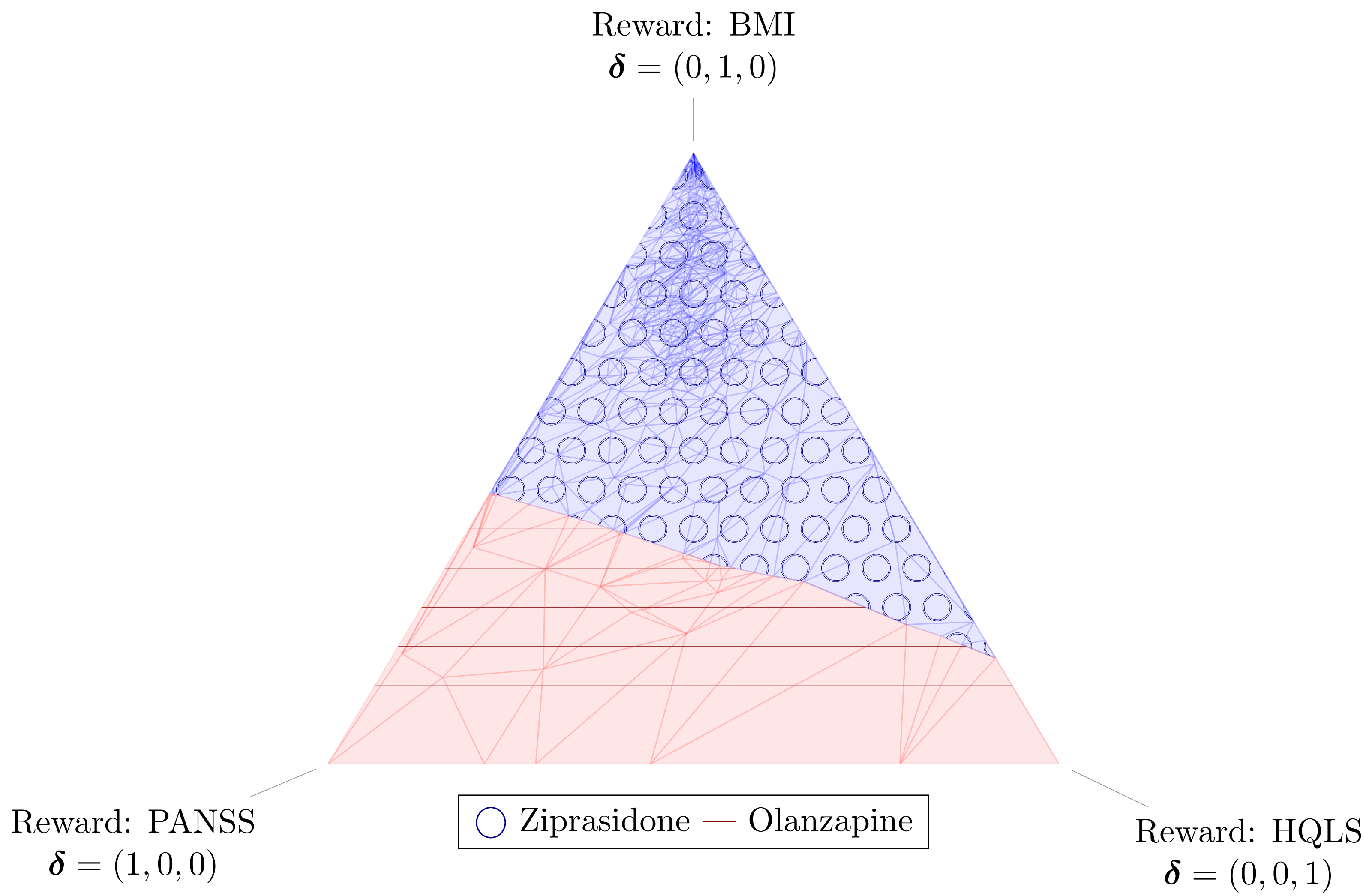
vs.

Body Mass Index

vs.

Heinrichs Quality of Life Scale

Phase 1



Preference Revealing

CS Challenges and Solutions

Value function/policy now a function of state **and** preference

Value functions **not convex** in preference, thus related methods for POMDPs do not apply

Computational geometry enables analysis of large, short-horizon trials

Multi-outcome Screening

1. Elicit "**clinically meaningful difference**" for each outcome
2. **Screen out** treatments that are "definitely bad"
3. Recommend the **set** of remaining treatments

Multi-outcome Screening

Suppose two* different rewards are important for decision making:

$r_{[1]}, r_{[2]}$

Screen out a treatment if another treatment is much worse for one reward and not much better for the other reward.

Do not screen if

1) treatments are not much different or
2) one treatment is much worse for one reward but much better for the other

Output: Set containing one or both treatments, possibly with a reason if both are included.

Multi-outcome Screening

Benefits

No notion of preference required

Suggests a **set** rather than imposing a single recommended treatment

Information still **individualized** through patient state

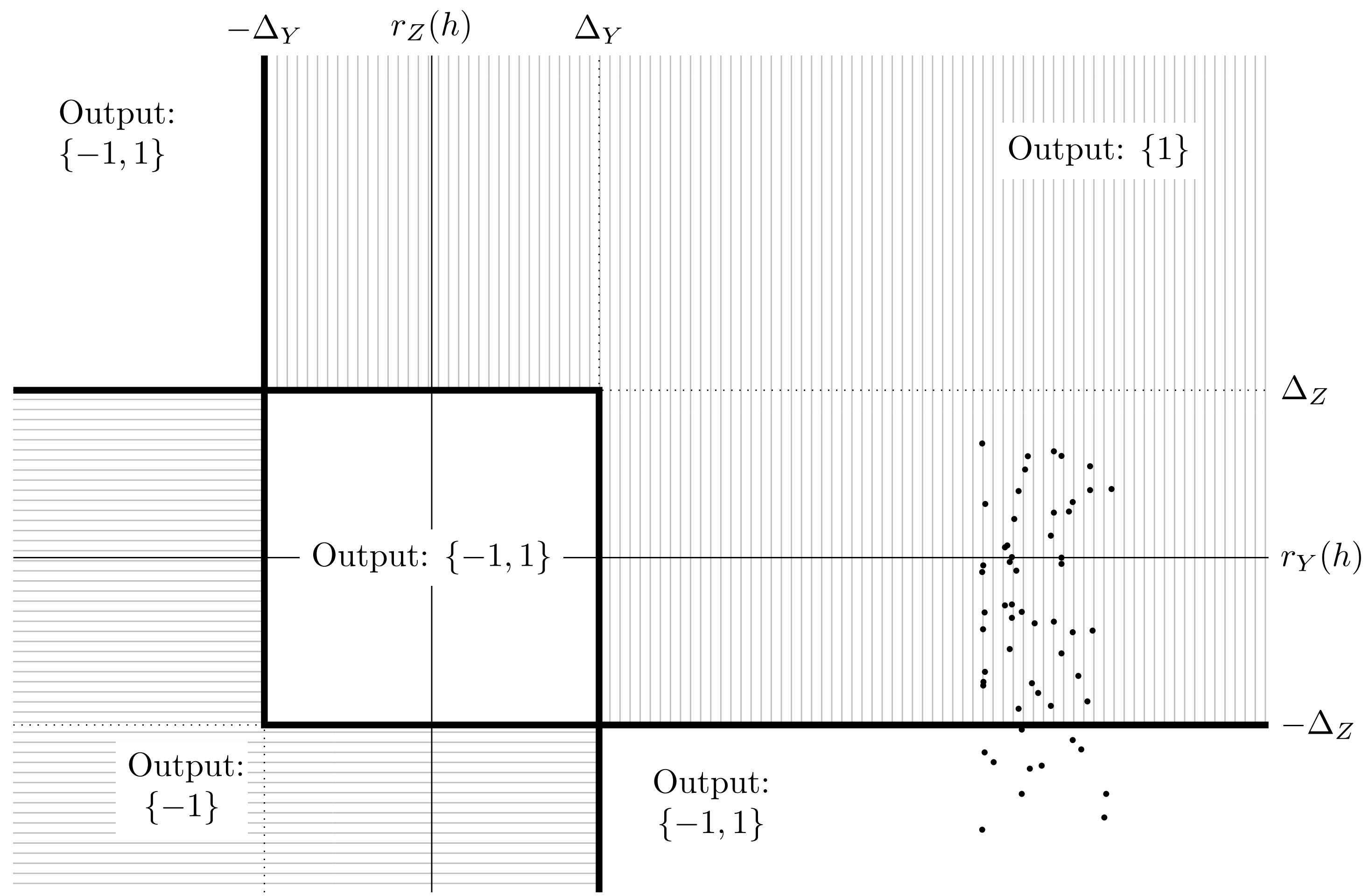
Treatments with bad evidence are **explicitly screened**

Screening **criterion is intuitive**

Positive And Negative Syndrome Scale
and
Body Mass Index

Phase 2 Efficacy

Y: PANSS, *Z*: BMI
-1: Not Clozapine, 1: Clozapine



Multi-outcome Screening

CS Challenges and Solutions

Lack of a unique policy means **dynamic programming** (e.g. Q-learning) **no longer works**

Must **consider all policies the user might follow** in future

Restriction to policies that 1) follow recommendations and 2) are "not too complex" **makes computation feasible**

Wrap-up

- Autonomous Agent model is for decision making; we want decision support.
- Part of good decision support is acknowledging different preferences
- Questions:
  - How can we add uncertainty information?
  - What about preferences changing over time?
  - What is the best way to convey information in a deployed application?
- Where else could this idea be useful?

References

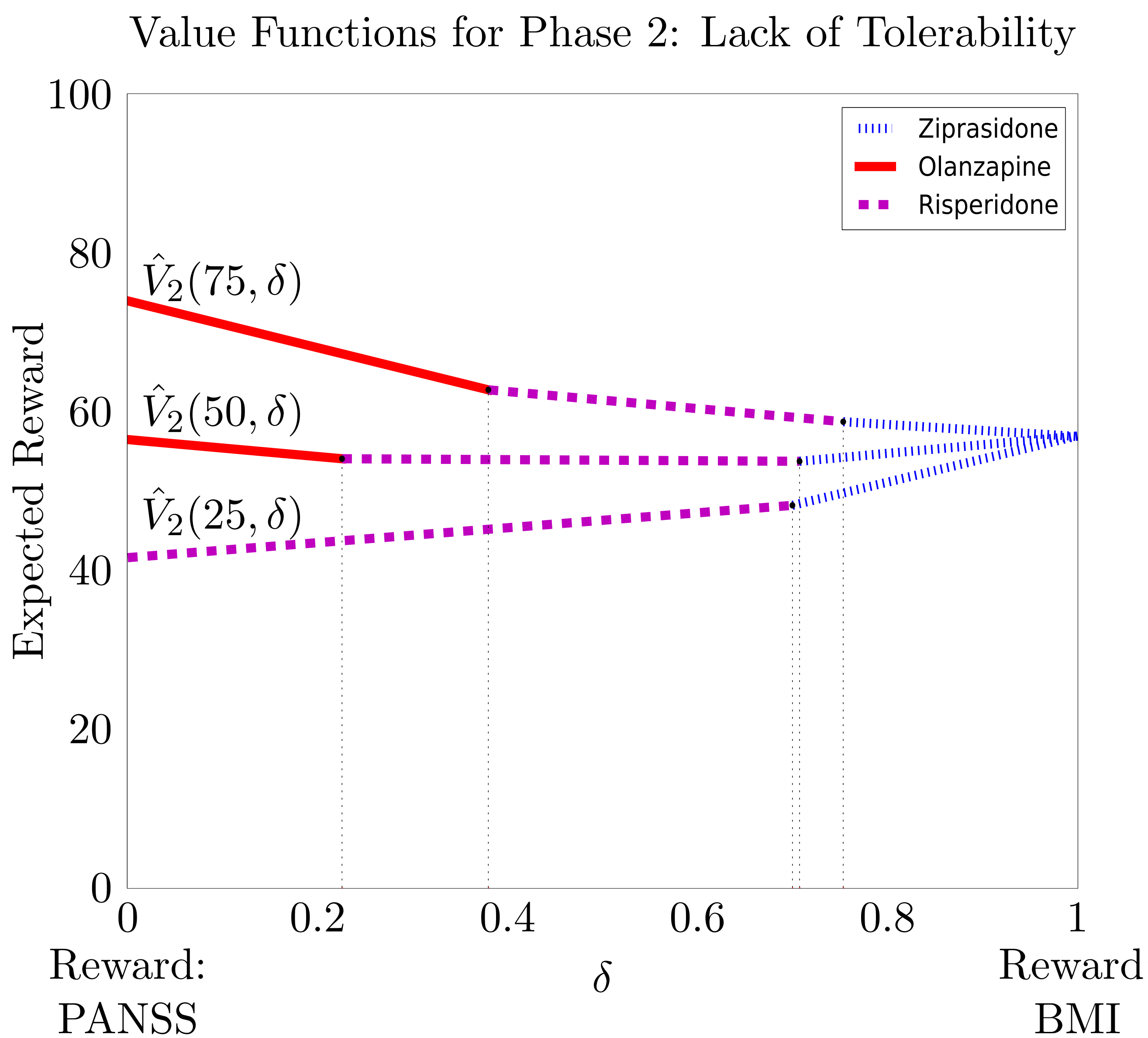
- Daniel J. Lizotte, Michael Bowling, and Susan A. Murphy. **Efficient Reinforcement Learning with Multiple Reward Functions for Randomized Clinical Trial Analysis**. Proc. ICML, 2010.
- Daniel J. Lizotte, Michael Bowling, and Susan A. Murphy. **Linear Fitted-Q Iteration with Multiple Reward Functions**. Accepted to Journal of Machine Learning Research.
- Eric B. Laber, Daniel J. Lizotte, Bradley Ferguson. **Set-valued dynamic treatment regimes for competing outcomes**. arXiv.

Positive And Negative Syndrome Scale

vs.

Body Mass Index

Phase 2 Tolerability

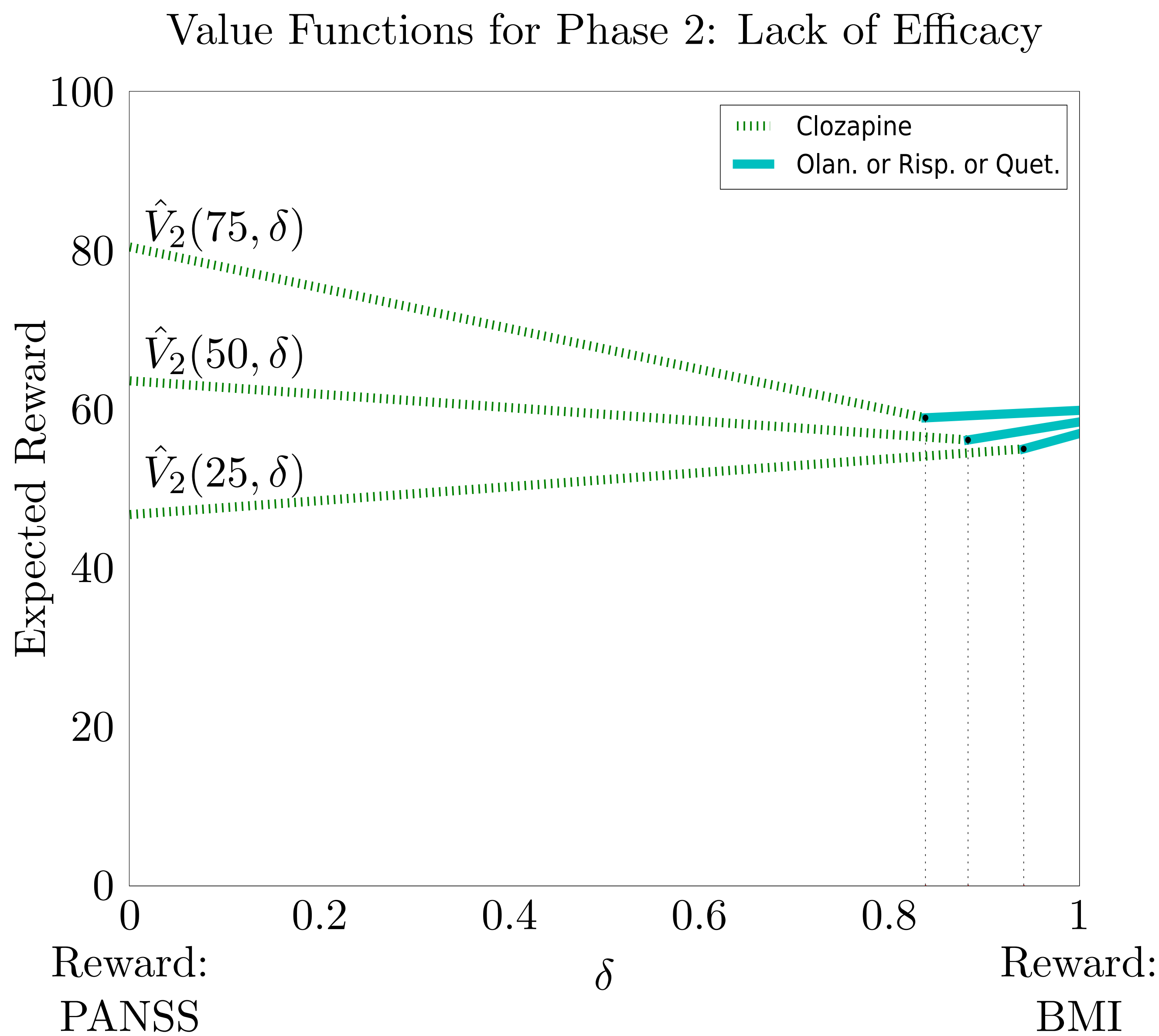


Positive And Negative Syndrome Scale

vs.

Body Mass Index

Phase 2 Efficacy



Positive And Negative Syndrome Scale

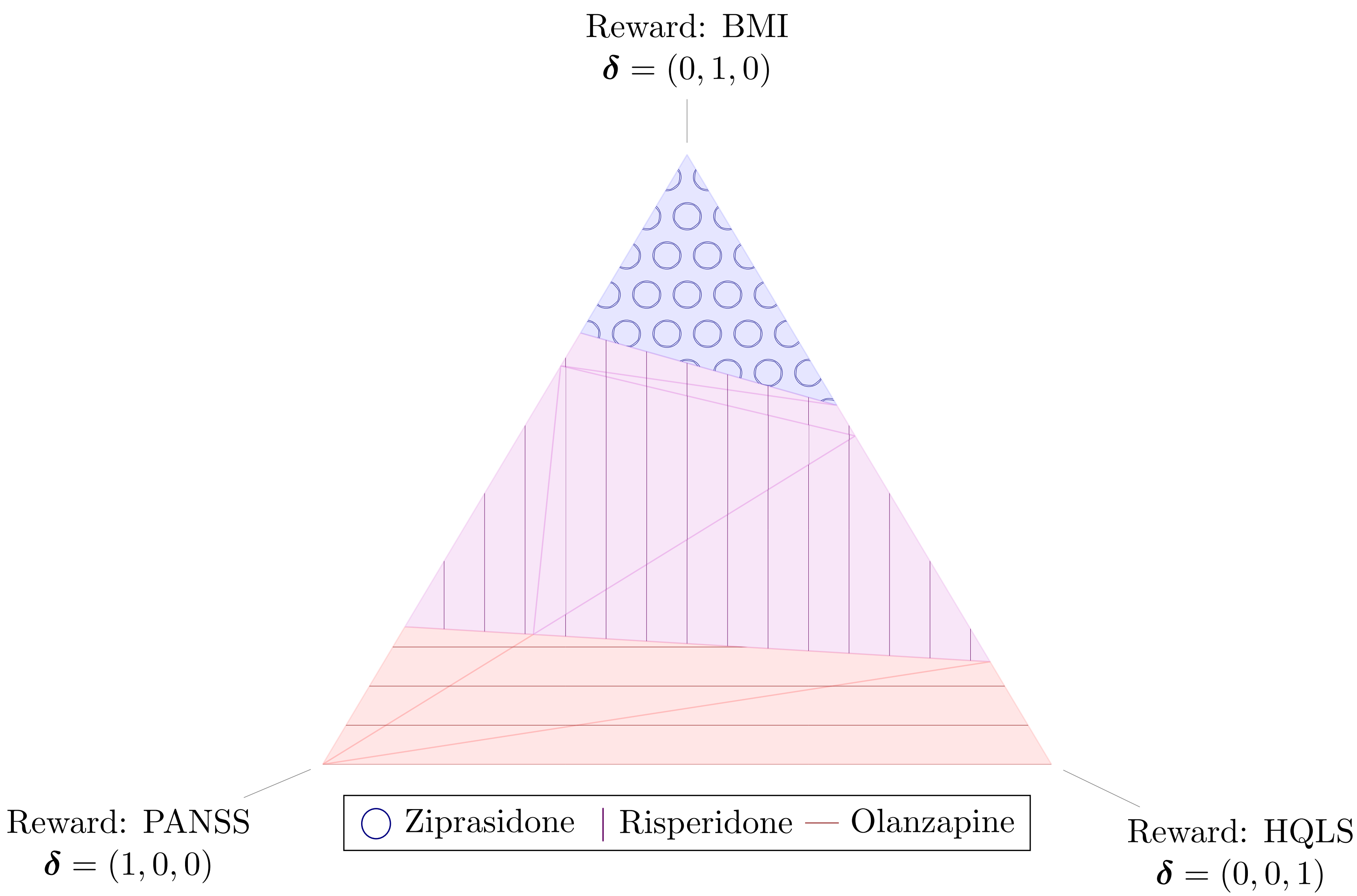
vs.

Body Mass Index

vs.

Heinrichs Quality of Life Scale

Phase 2 Tolerability



Positive And Negative Syndrome Scale

vs.

Body Mass Index

vs.

Heinrichs Quality of Life Scale

Phase 2 Efficacy