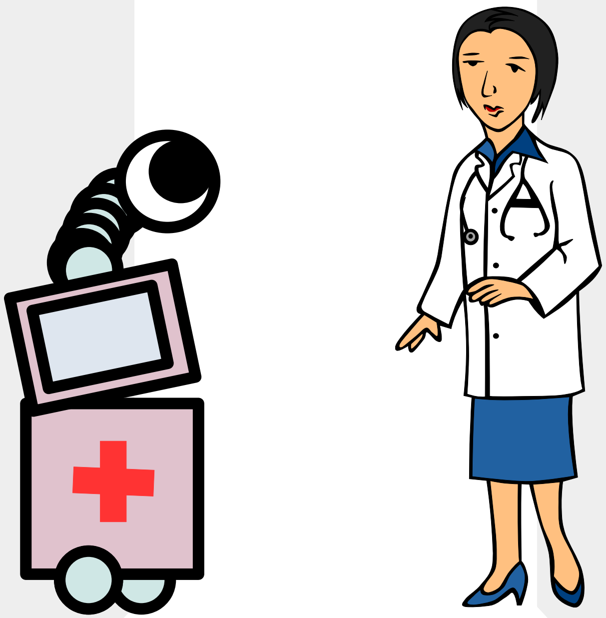
Towards Better Healthcare with AIs Made for Human Validation

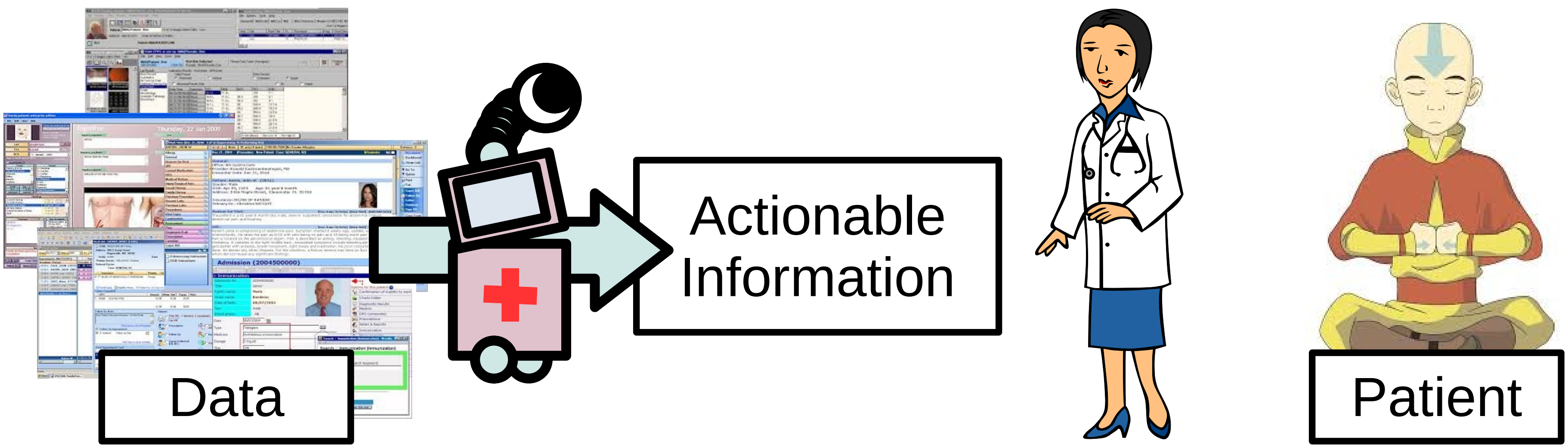
Finale Doshi-Velez



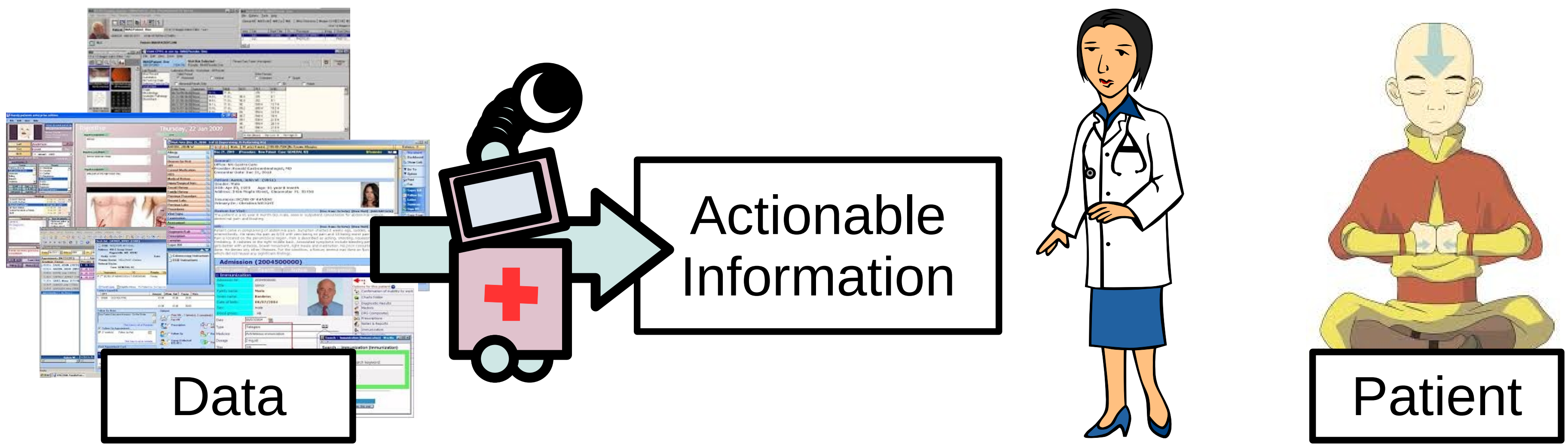
In collaboration with the absolutely wonderful

DtAK and DtAK alums: Weiwei Pan, Sonali Parbhoo, Melanie Pradier, Joe Futoma, Michael Hughes, Madhi Pakdaman, Ike Lage, Andrew Ross, Yaniv Yacoby, Jiayu Yao, Beau Coker, Anna Li, Sarah Rathnam, Abhishek Sharma, Eura Shin, Omer Gottesman, Muhammad Arjumand Masood; **Collaborators**: Roy Perlis, Tom McCoy, Taylor Killian, Soumya Ghosh, Xuefeng Peng, David Wihl, Yi Ding, Liwei Lehman, Matthieu Komorowski, Aldo Faisal, David Sontag, Fredrik Johansson, Leo Celi, Aniruddh Raghu, Yao Liu, Emma Brunskill, Sam Gershman, Been Kim, Menaka Narayanan, Emily Chen, Jeffrey He, Ofra Amir, and the CS282 2017; **Admins**: Meg Hastings, Michaela Kapp, Jenny Mileski, Ashley Bens, Annalee Mendez, Jill Sussery, Jasmin Ware, Joanne Bourgeois… and **many, many more** supporters and students at SEAS and beyond!

Our lab: Novel AI to Support Human Decision-Making in Health

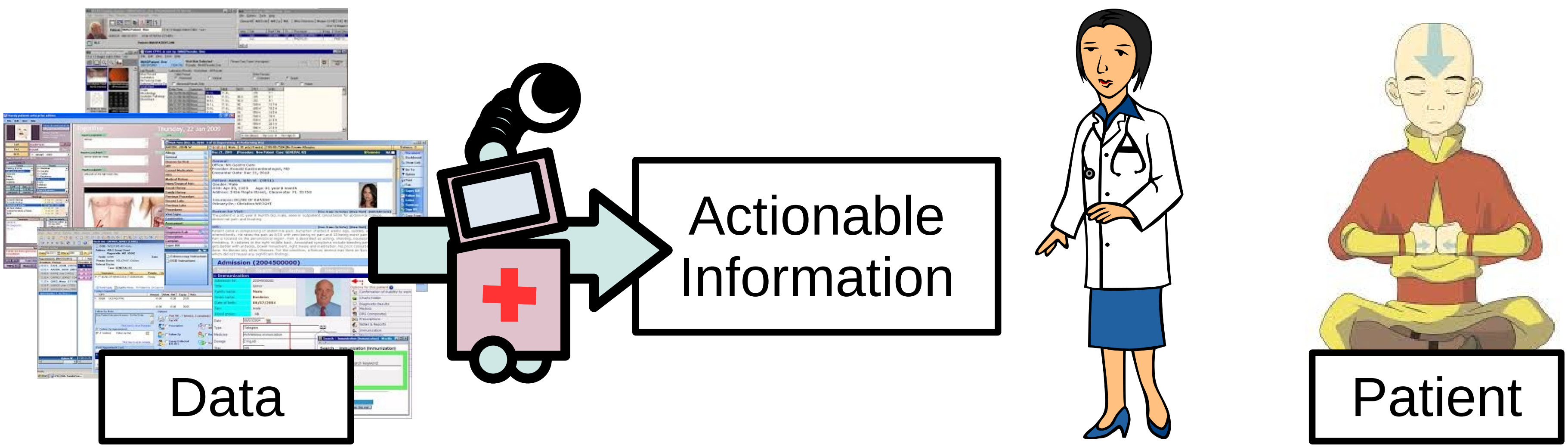


Our lab: Novel AI to Support Human Decision-Making in Health



Limited fields; entry errors; patients come when sick; clinician goals unknown

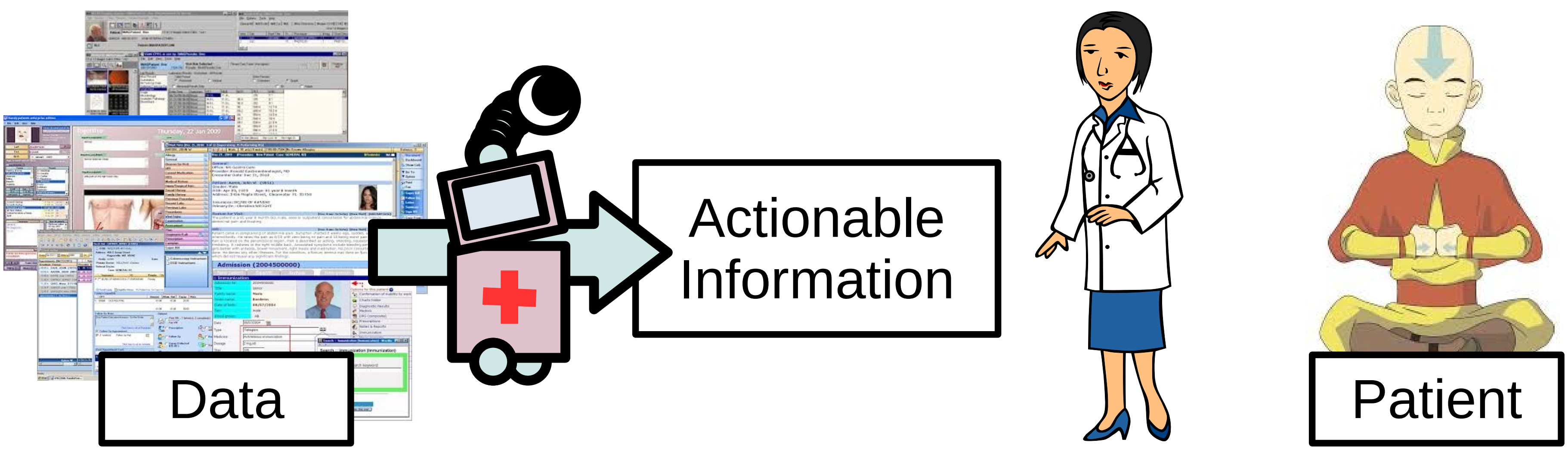
Our lab: Novel AI to Support Human Decision-Making in Health



Limited fields; entry errors; patients come when sick; clinician goals unknown

What, how to share so humans make good decisions in risky situations

Our lab: Novel AI to Support Human Decision-Making in Health

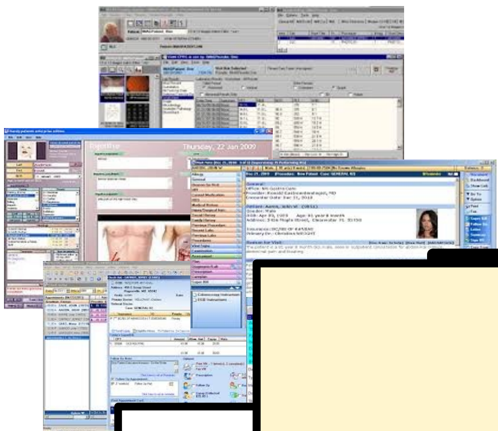


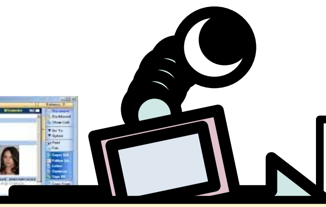
Limited fields; entry errors; patients come when sick; clinician goals unknown

What, how to share so humans make good decisions in risky situations

Not all relevant info recorded

Our lab: Novel AI to Support Human Decision-Making in Health





Actionable

Da

ient

Let’s see this process
with an example!

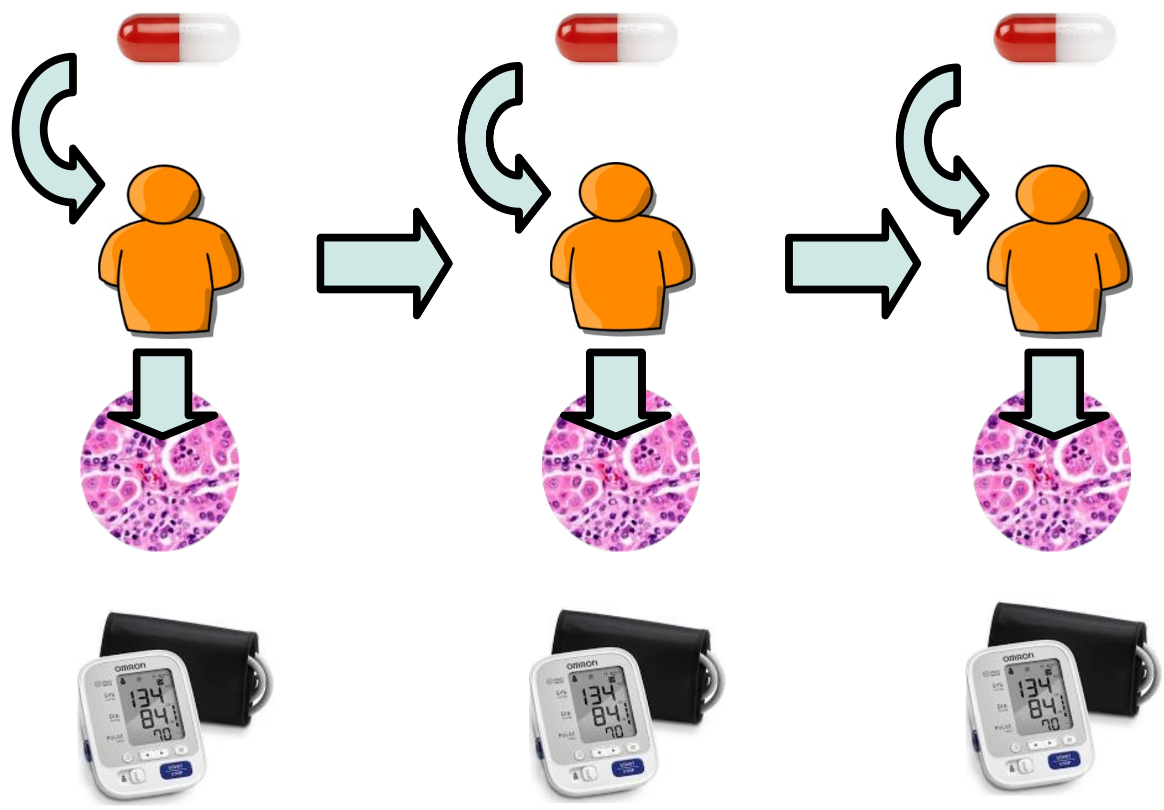
Limited fie
errors; pa
when sick; clinician
goals unknown

good decisions in
risky situations

relevant
corded

Example: Optimizing HIV treatments

Goal: Manage HIV, avoid resistance



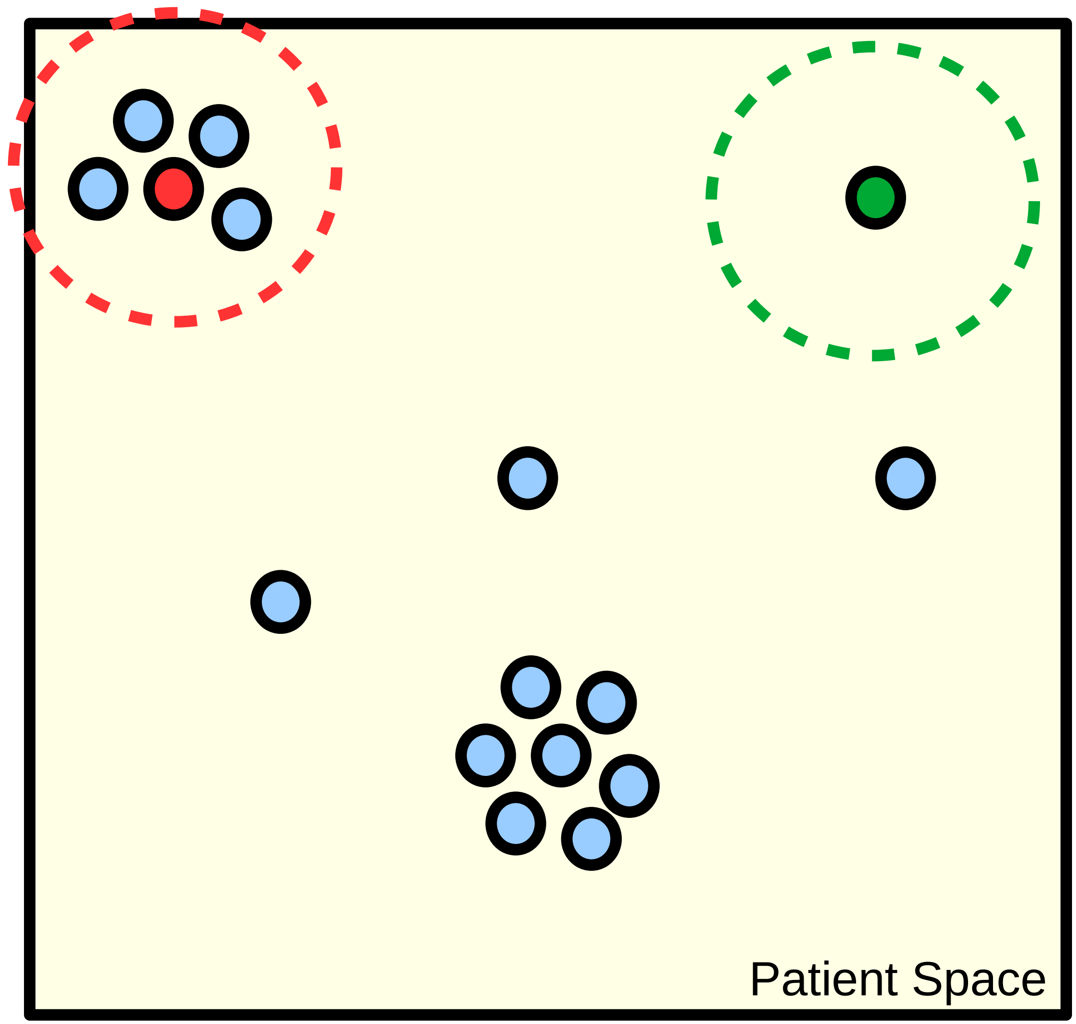
Example: Optimizing HIV treatments

Goal: Manage HIV, avoid resistance

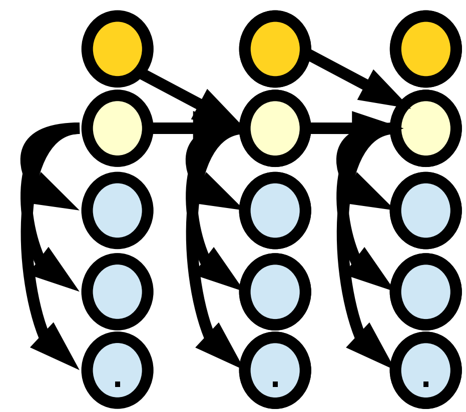
Step 1: How should we **model** complex disease processes?

Our **modeling insight**: Combine models with complementary strengths

Patients in clusters may be best modeled by their neighbors



Patients without neighbors may be better modeled with a model



Step 2: **Optimize** Combination Policy

h(

Neighbor Action

Model Action

Patient, Cohort Statistics

) =

Actual Action

Step 3: **Validate** (It works!!)

- 32,960 patients from EU Resist Database; hold out 3,000 for testing.
- Observations: CD4s, viral loads, mutations
- Actions: 312 drug combos (from 20 drugs)

Approach	DR Reward
Random Policy	-7.31 ± 3.72
Neighbor Policy	9.35 ± 2.61
Model-Based Policy	3.37 ± 2.15
Policy-Mixture Policy	11.52 ± 1.31
Model-Mixture Policy	12.47 ± 1.38

*Mixture chooses POMDP about 30% of the time.

Parbhoo ML4HWS (best paper) 2016; Parbhoo et al., AMIA 2017; Parbhoo et al. PloS Med 2018

Extension: Transfer from EU cohort to South African cohort

We identified when policies in well-curated EU cohorts to could help patients in less-well curated SA cohorts.

Type	Method	DR	IS	WIS
Behaviour Policy	5.02 ± 1.18			
Local	Kernel	3.56 ± 1.42	1.27 ± 1.14	1.80 ± 1.07
	CEIB	3.29 ± 1.13	3.80 ± 2.41	3.76 ± 2.19
Transfer	Kernel	4.17 ± 1.4	4.18 ± 1.20	4.16 ± 1.71
	CEIB	6.29 ± 0.14	5.17 ± 0.38	5.27 ± 0.29
	Mixture-of-Experts	5.28 ± 0.37	3.42 ± 1.39	4.81 ± 1.25
Local + Transfer	**Ours**	$\mathbf{8.96 \pm 0.39}$	$\mathbf{10.64 \pm 1.2}$	$\mathbf{10.62 \pm 1.67}$

Parbhoo et al., MLHC 2020

Extension: hypotension management

From a cohort of 15,415 ICU patients, optimized vasopressor and fluid use to minimize 30-day mortality.

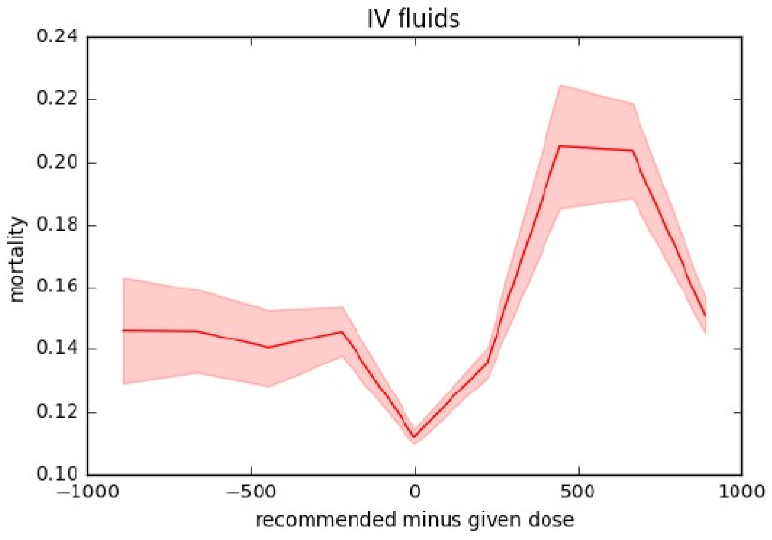
	Physician	Kernel	DQN	MoE_{V_d,Q_d}	MoE_{V_b,Q_b}
non-recurrent encoded	3.76	3.73	4.06	3.93	4.31
recurrent encoded	3.76	4.46	4.23	5.03	5.72

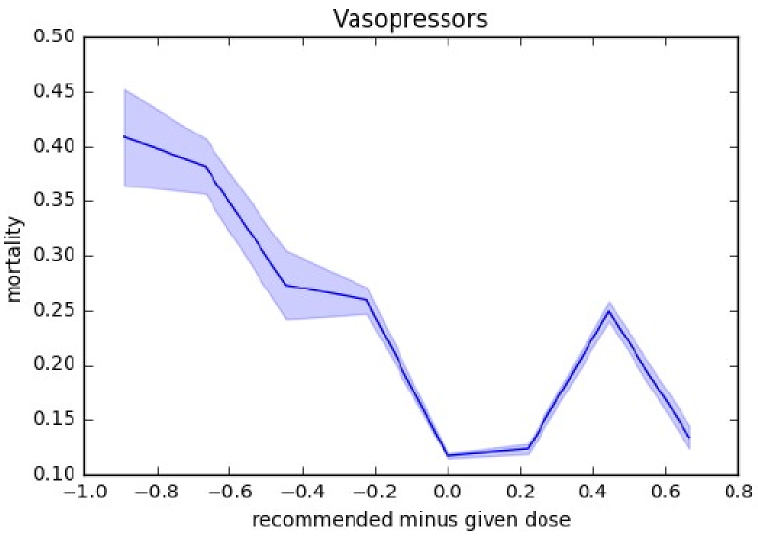
CS282 2017; Gottesman et al. Nat Med 2019

Extension: hypotension management

From a cohort of 15,415 ICU patients, optimized vasopressor and fluid use to minimize 30-day mortality.

	Physician	Kernel	DQN	MoE_{V_d,Q_d}	MoE_{V_b,Q_b}
non-recurrent encoded	3.76	3.73	4.06	3.93	4.31
recurrent encoded	3.76	4.46	4.23	5.03	5.72



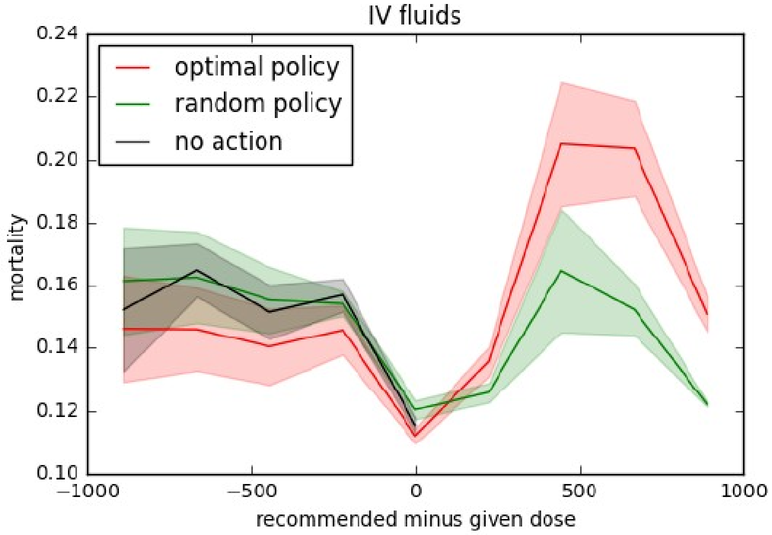


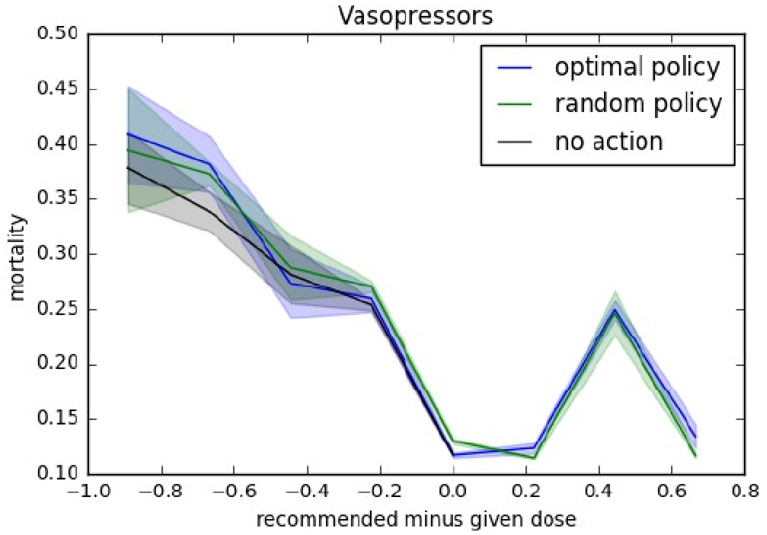
CS282 2017; Gottesman et al. Nat Med 2019

Extension: hypotension management

From a cohort of 15,415 ICU patients, optimized vasopressor and fluid use to minimize 30-day mortality.

	Physician	Kernel	DQN	MoE_{V_d,Q_d}	MoE_{V_b,Q_b}
non-recurrent encoded	3.76	3.73	4.06	3.93	4.31
recurrent encoded	3.76	4.46	4.23	5.03	5.72





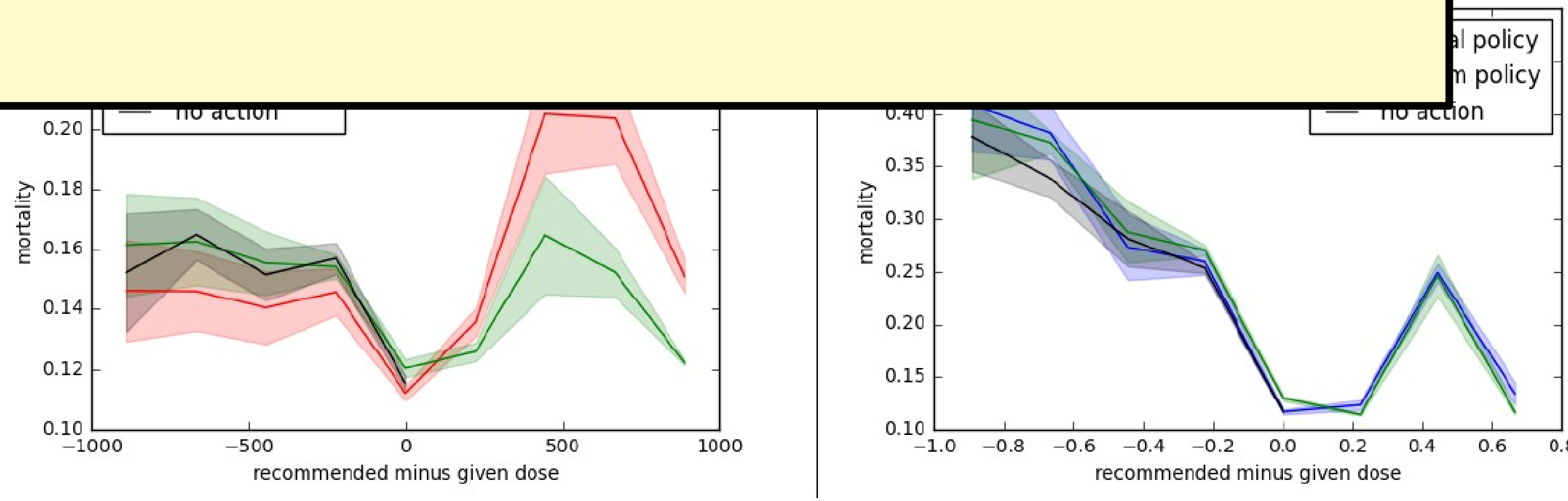
CS282 2017; Gottesman et al. Nat Med 2019

Extension: hypotension management

From a cohort of 15,415 ICU patients, optimized vasopressor and fluid use to minimize 30-day mortality.

	MoE_{V_b,Q_b}
n	4.31
	5.72

All stats will have limits.



CS282 2017; Gottesman et al. Nat Med 2019

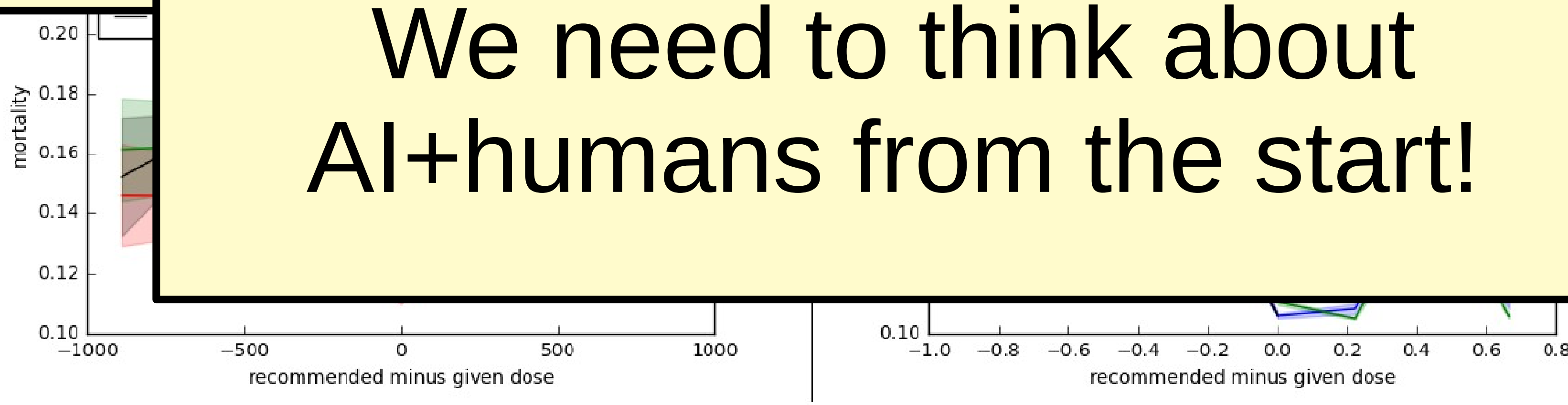
Extension: hypotension management

From a cohort of 15,415 ICU patients, optimized vasopressor and fluid use to minimize 30-day mortality.

	MoE_{V_b,Q_b}
n	4.31
	5.72

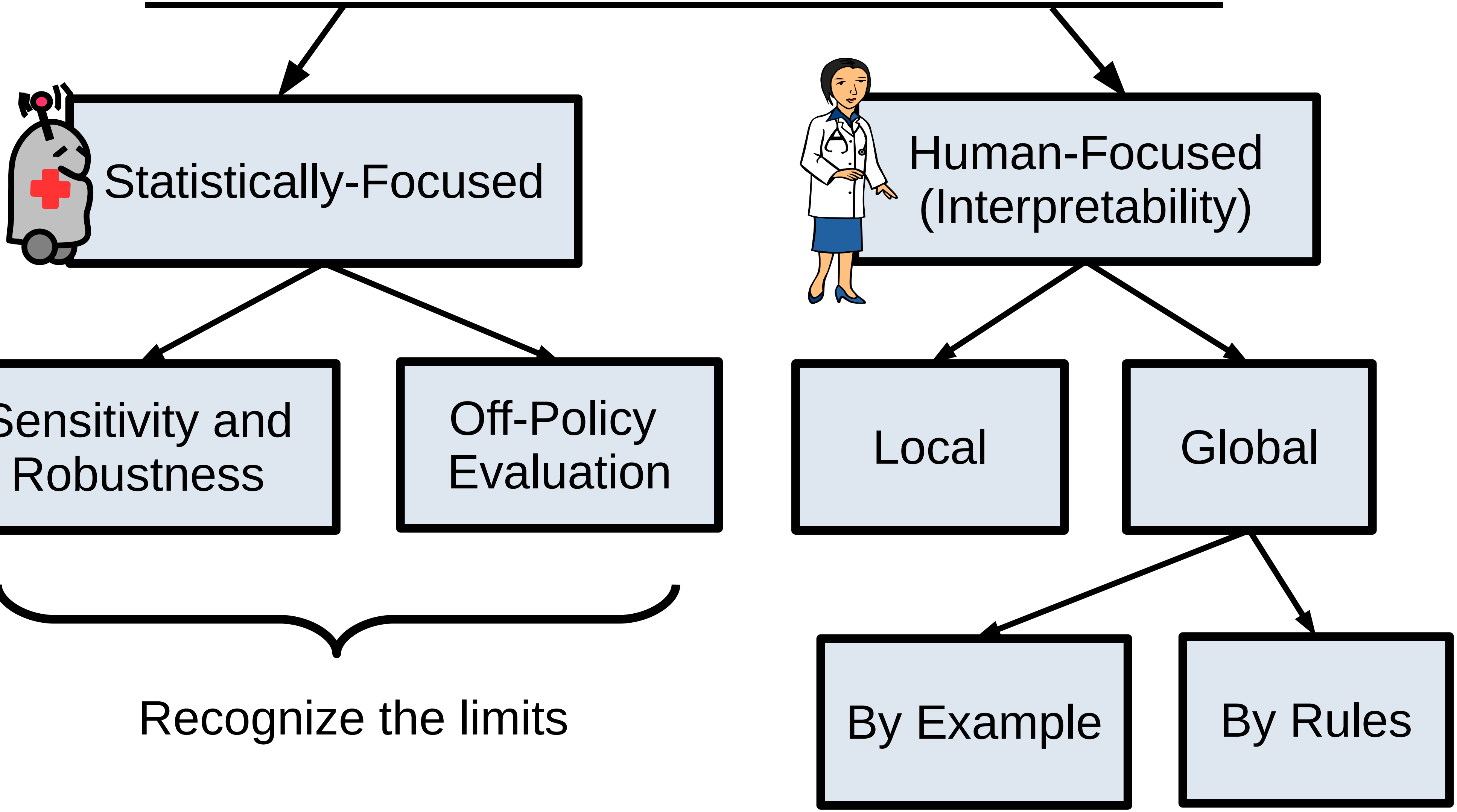
All stats will have limits.

We need to think about AI+humans from the start!

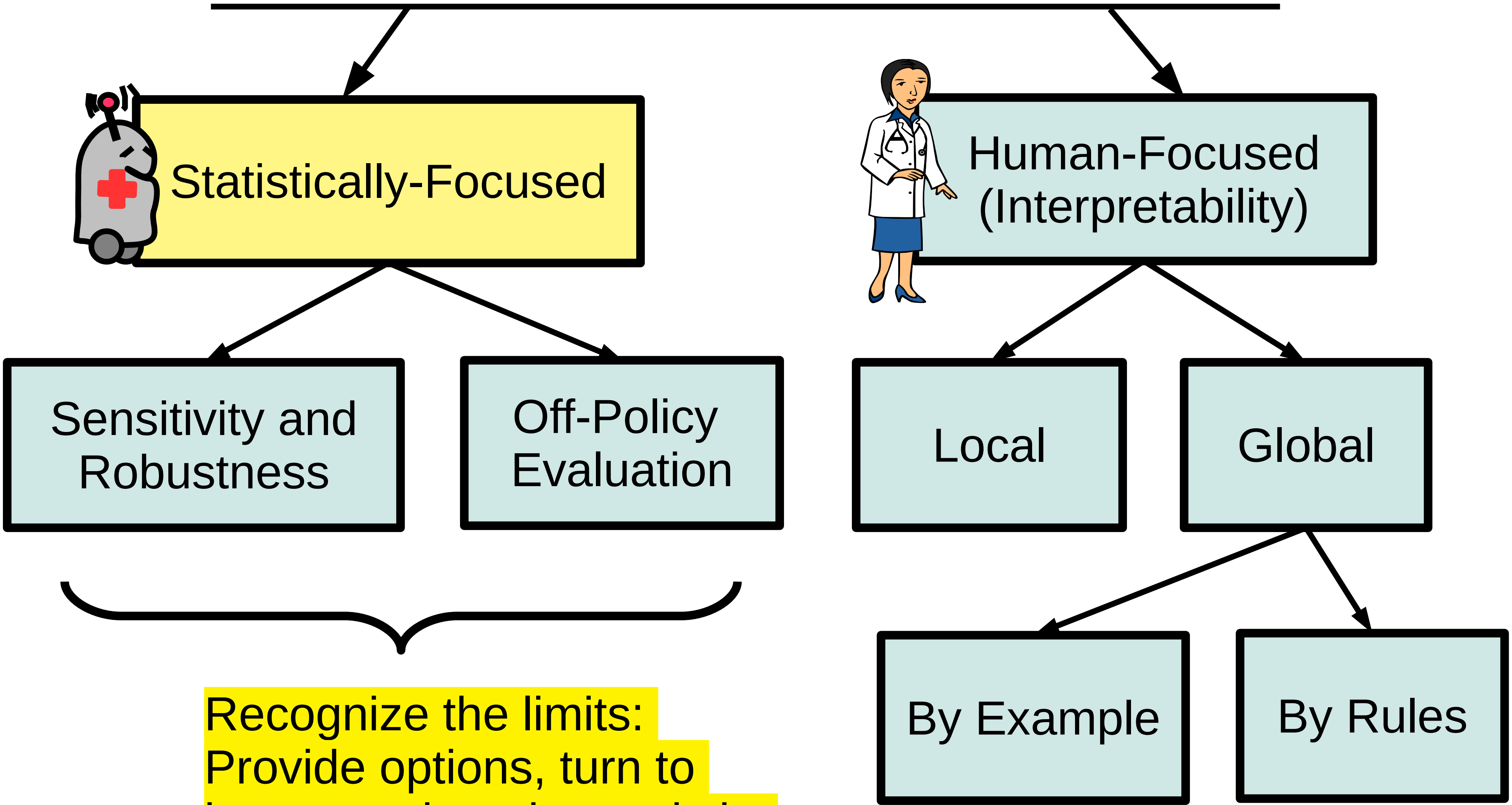


CS282 2017; Gottesman et al. Nat Med 2019

Batch Validation Roadmap

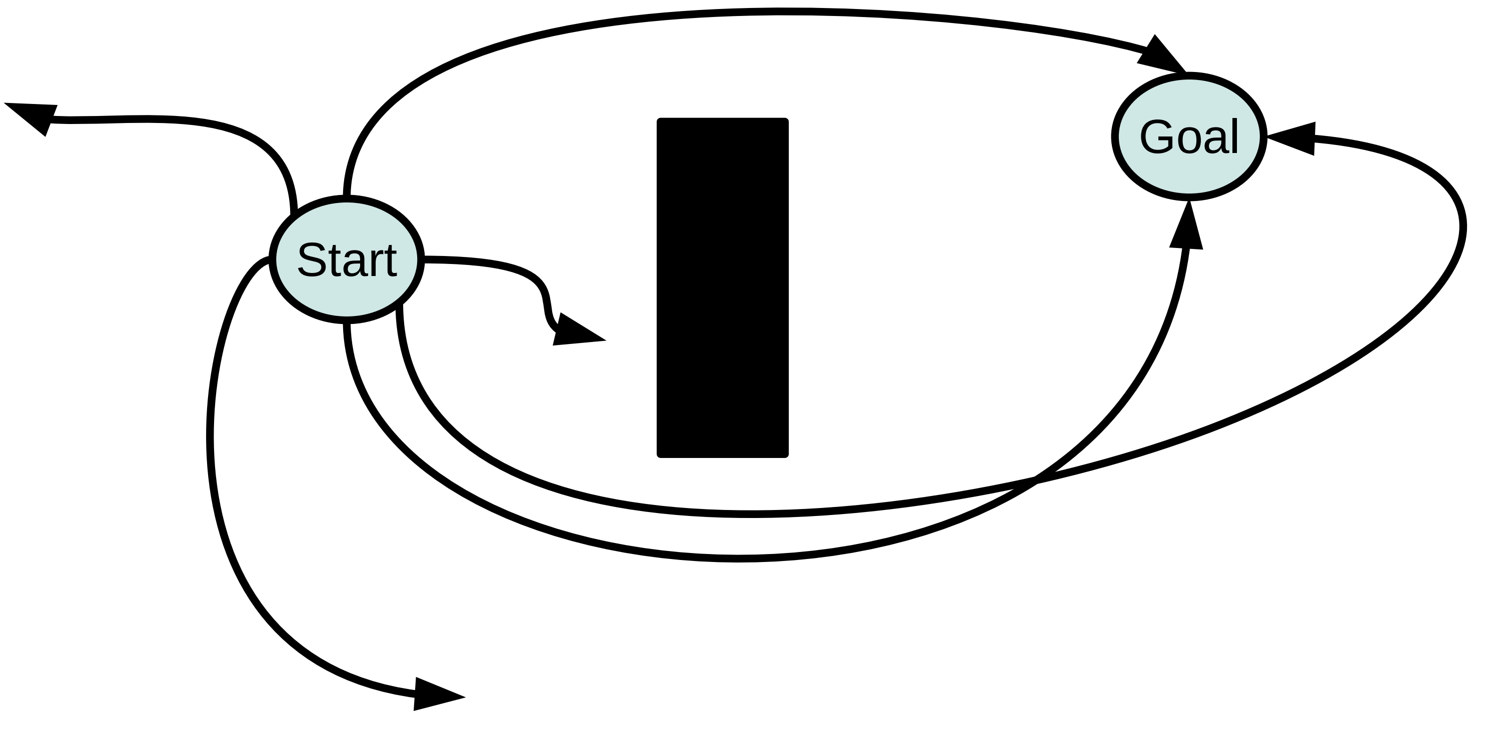


Batch Validation Roadmap



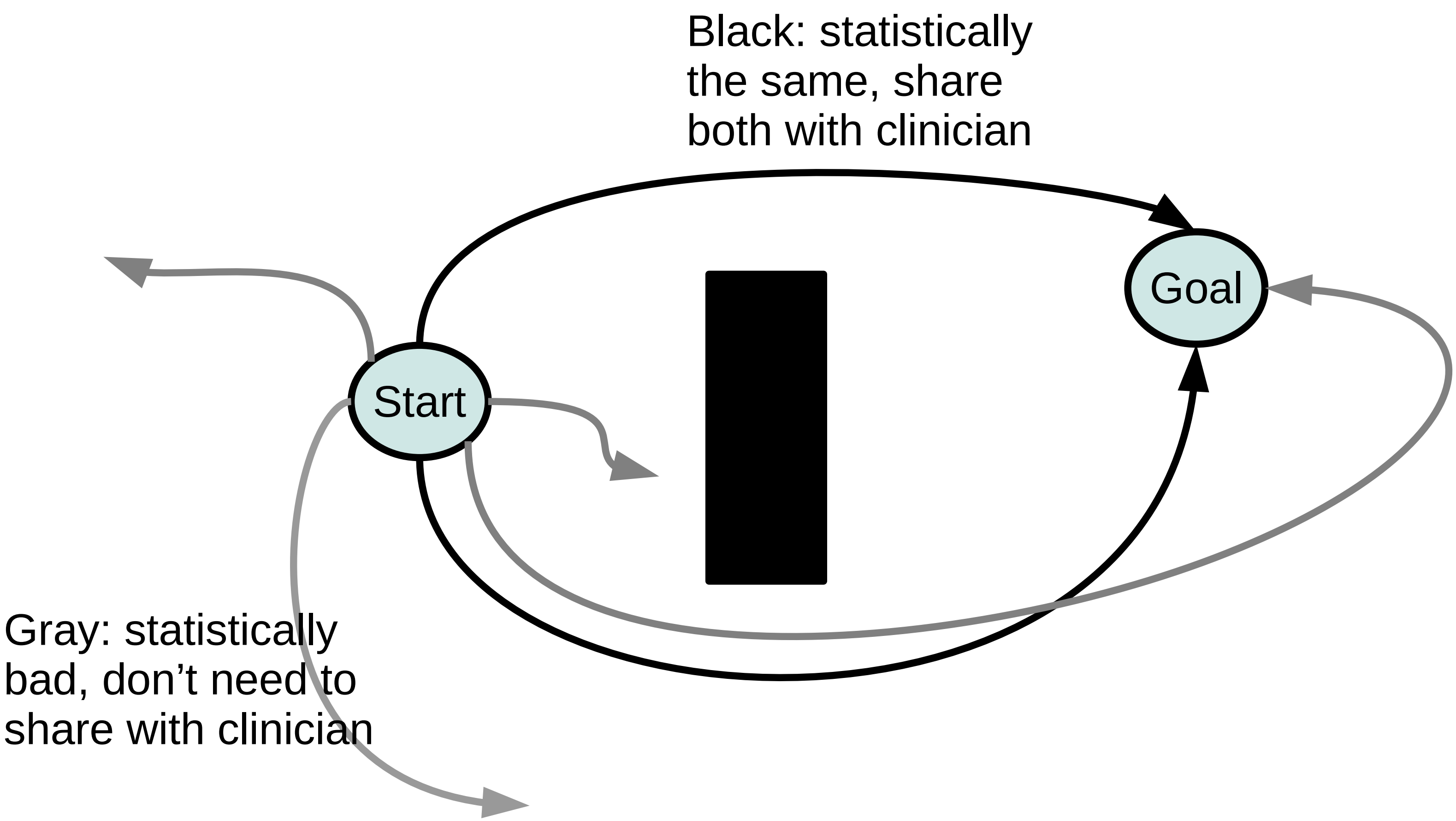
Recognize the limits: Provide options, turn to humans when the statistics cannot tell you more.

Provide options when the statistics cannot tell you more



(Masood et al, IJCAI 2019; Futoma et al. AMIA 2020)

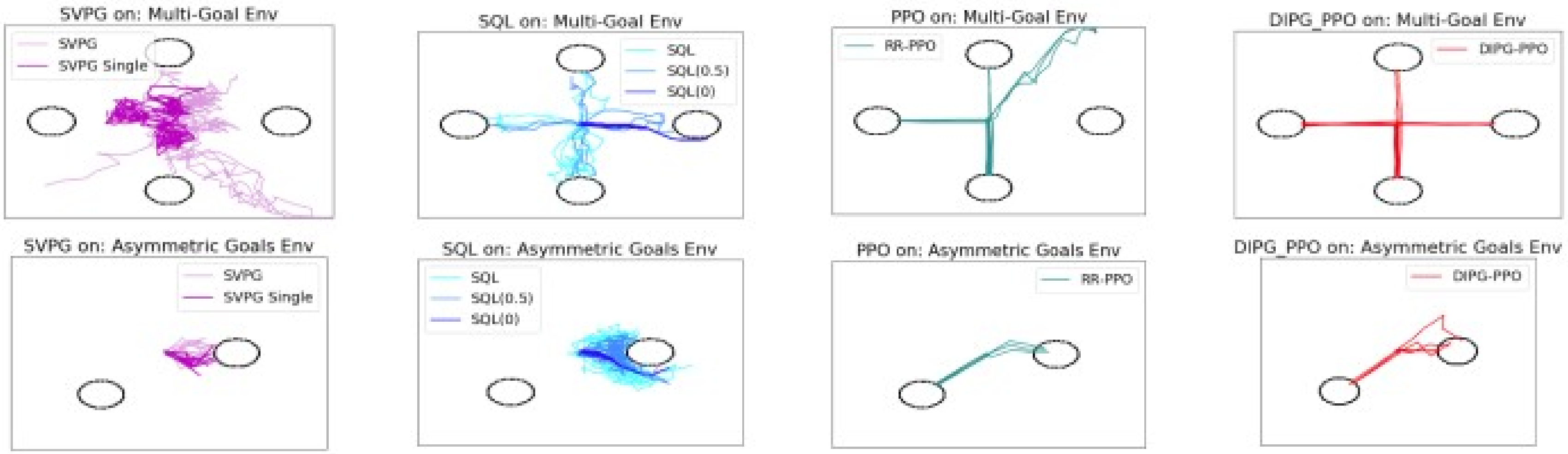
Provide options when the statistics cannot tell you more



(Masood et al, IJCAI 2019; Futoma et al. AMIA 2020)

Displaying Diverse Alternatives

If policies can't be statistically differentiated, give plausible alternatives.



(Masood et al, IJCAI 2019)

Providing Options

If we can't identify the optimal strategy, suggest some reasonable ones

$$\Pi^* = \text{argmin}\ \mathcal{L}_Q(\pi_{beh}, \Pi) + \lambda \mathcal{L}_D(\Pi), \ \ \text{s.t.}\ \pi = \text{safe}_{\pi_{beh}}(\pi), \forall \pi \in \Pi$$

A collection of policies π from some class.

Quality: How good is each policy?

Diversity: How different are the policies? (Use KL)

Safety: Forbid rare and unseen actions (hard constraint)

(Masood et al, IJCAI 2019; Futoma et al. AMIA 2020)

Providing Options

If we can't identify the optimal strategy, suggest some reasonable ones (Futoma et al. 2020)

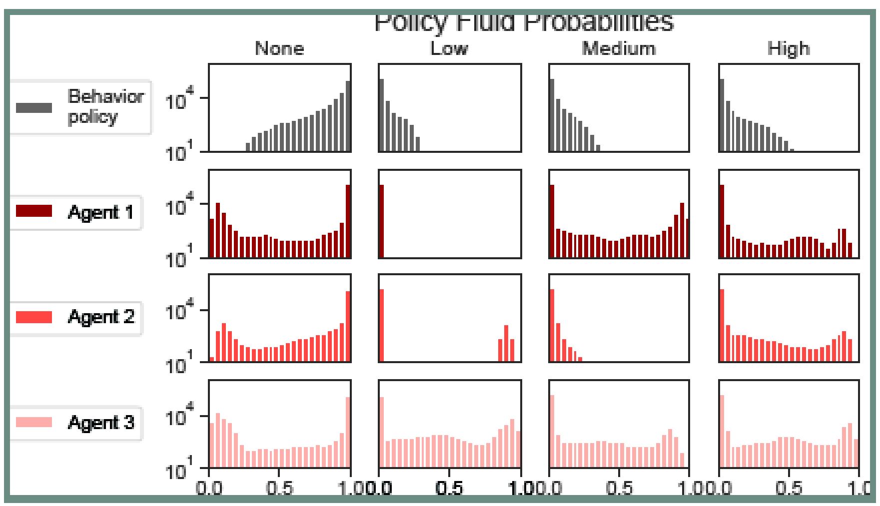
	Setting			Behavior: 37.90 Quantitative Metrics					
	Diversity Term?	Safety Mask?	Quality Term?	# Kept Agents	OPE Est. Value	ESS	Difference from Beh.	Diversity	% Rare Actions
Diverse and Safe	Yes	Yes	Yes	3	35.43 ± 1.45	222 ± 102	0.58 ± 0.06	2.05 ± 0.23	0.0 ± 0.0
	Yes	Yes	No	4	35.86 ± 1.51	311 ± 181	1.39 ± 0.47	3.27 ± 0.00	0.0 ± 0.0
Diverse, not Safe	Yes	No	Yes	2	41.74 ± 0.36	235 ± 146	0.92 ± 0.32	2.90 ± 0.00	18.3 ± 7.8
	Yes	No	No	0	-	-	-	-	-
Safe, not Diverse	No	Yes	Yes	4	36.74 ± 0.08	284 ± 27	0.06 ± 0.00	0.00 ± 0.00	0.0 ± 0.0
Not Safe or Diverse	No	No	Yes	0	-	-	-	-	-

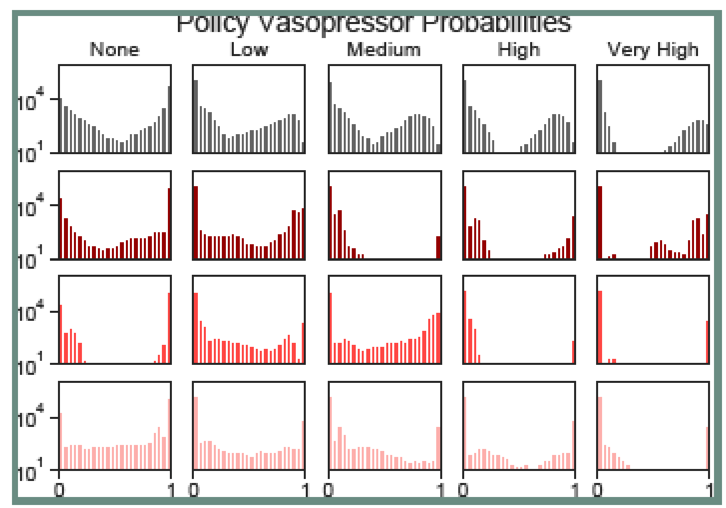
(Masood et al, IJCAI 2019; Futoma et al. AMIA 2020)

Providing Options

If we can't identify the optimal strategy, suggest some reasonable ones

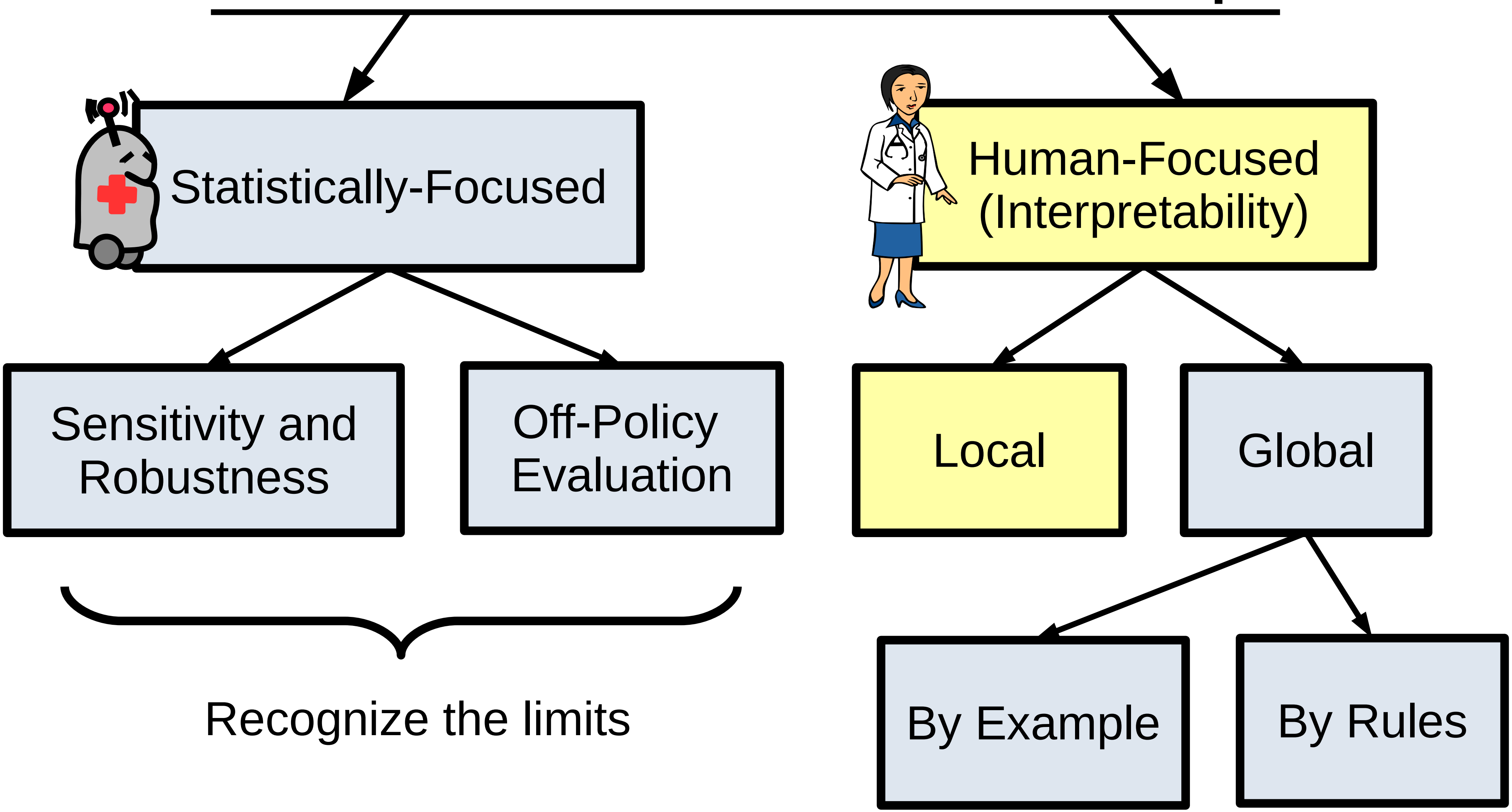
	Setting			Behavior: 37.90 Quantitative Metrics					
	Diversity Term?	Safety Mask?	Quality Term?	# Kept Agents	OPE Est. Value	ESS	Difference from Beh.	Diversity	% Rare Actions
Diverse and Safe	Yes	Yes	Yes	3	35.43 ± 1.45	222 ± 102	0.58 ± 0.06	2.05 ± 0.23	0.0 ± 0.0
	Yes	Yes	No	4	35.86 ± 1.51	311 ± 181	1.39 ± 0.47	3.27 ± 0.00	0.0 ± 0.0
Diverse, not Safe	Yes	No	Yes	2	41.74 ± 0.36	235 ± 146	0.92 ± 0.32	2.90 ± 0.00	18.3 ± 7.8
	Yes	[illegible]	[illegible]	[illegible]					
Safe, not Diverse	No					4			
Not Safe or Diverse	No								





(Masood et al, IJCAI 2019; Futoma et al. AMIA 2020)

Batch Validation Roadmap



Finding Errors: Do Explanations Help?

Patient Details:
Susan is a 31 year old woman who is single and works part time. She has a history of diabetes, arrhythmia and hypertensive heart disease. She presents with 14 months of depressed mood. Current medications include amoxicillin, and prior treatment with Paroxetine was ineffective.

System.13 Recommendation: **DULOXETINE**

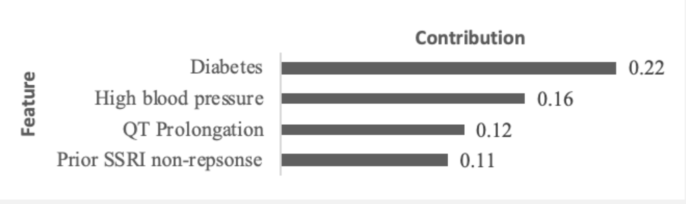
Top 5 therapies with highest probability for stability:

Therapy	Predicted Stability*	Predicted Dropout Risk**
Duloxetine	.80	.06
Fluoxetine	.68	.12
Citalopram	.67	.13
Escitalopram	.59	.16
Bupropion	.57	.20

**Stability: continued use of the same medication for at least 3 months*
***Dropout: early treatment discontinuation following prescription*

Why are these therapies being recommended?

The following **patient features** had the highest contributions to system.13's predictions:



Patient Details:
Thomas is a 38 year old man who is single and works full time. He has a history of diabetes, hypertensive heart disease, and arrhythmia. He presents with 10 months of depressed mood. Current medications include amoxicillin, and prior treatment with Paroxetine was ineffective.

System.16 Recommendation: **DULOXETINE**

Top 5 therapies with highest probability for stability:

Therapy	Predicted Stability*	Predicted Dropout Risk**
Duloxetine	.77	.04
Fluoxetine	.69	.07
Citalopram	.65	.07
Escitalopram	.65	.07
Bupropion	.51	.18

**Stability: continued use of the same medication for at least 3 months*
***Dropout: early treatment discontinuation following prescription*

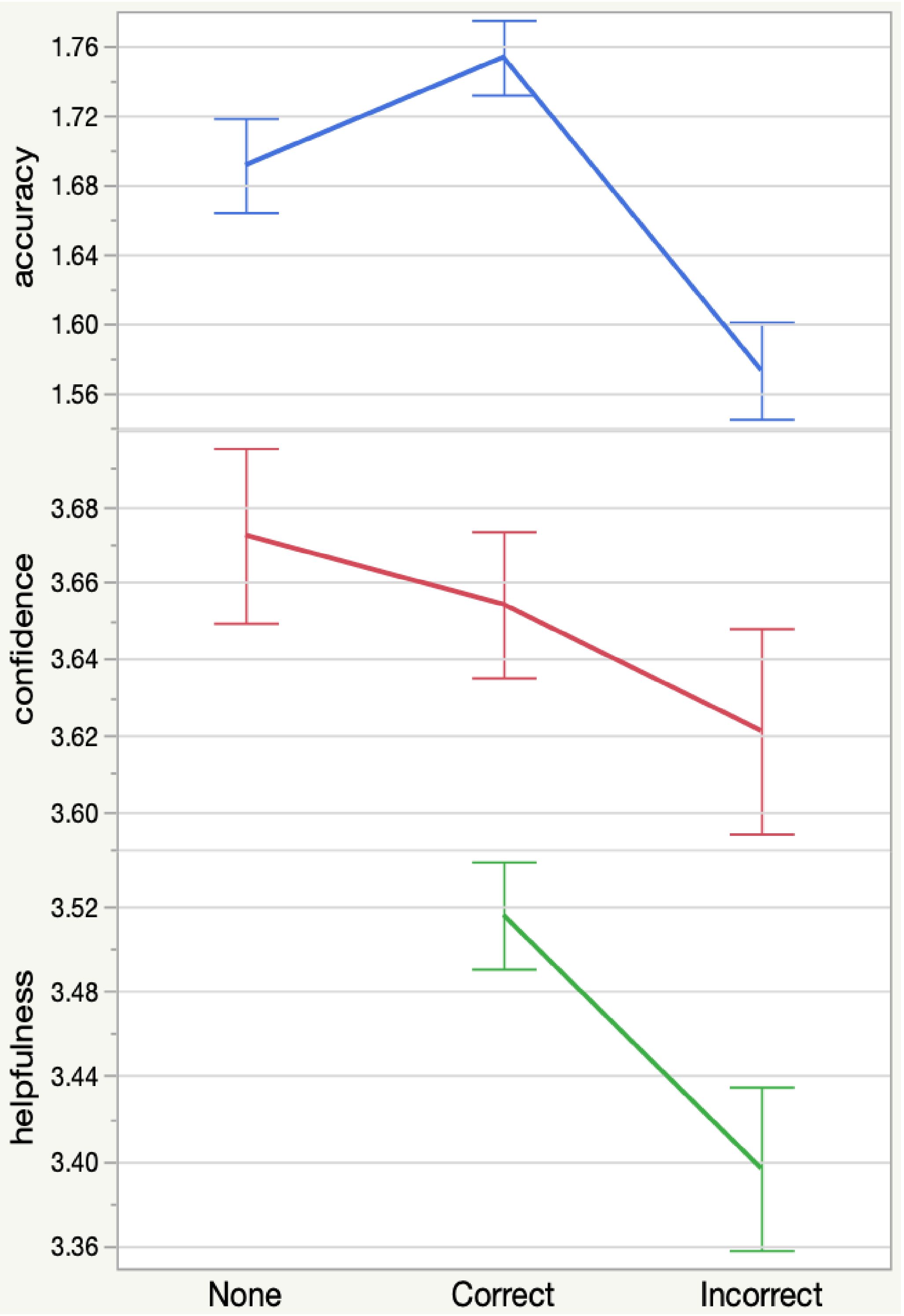
Why are these therapies being recommended?

The following **rules** had the highest contributions to system.16's predictions:

1. *If concern for QT prolongation, favor Sertraline, avoid Citalopram*
2. *If avoiding weight gain, favor weight loss, favor Bupropion, avoid Mirtazapine*
3. *If concern for increased blood pressure, avoid SNRI's*
4. *If lack of response to Paroxetine, avoid SSRI's*

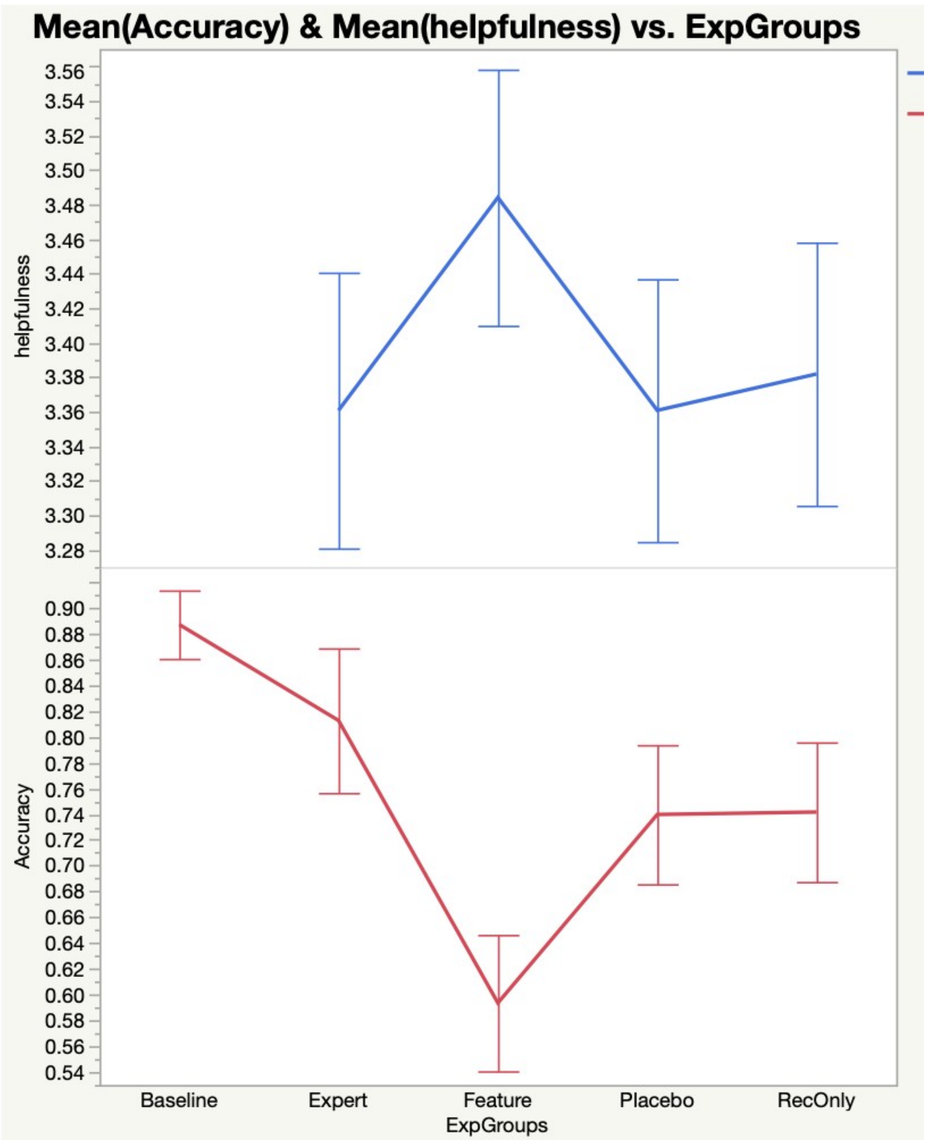
Results

- Accuracy increases if the recommendation is correct, decreases if not correct.
- Some awareness of helpfulness w.r.t. the correctness of the explanation.

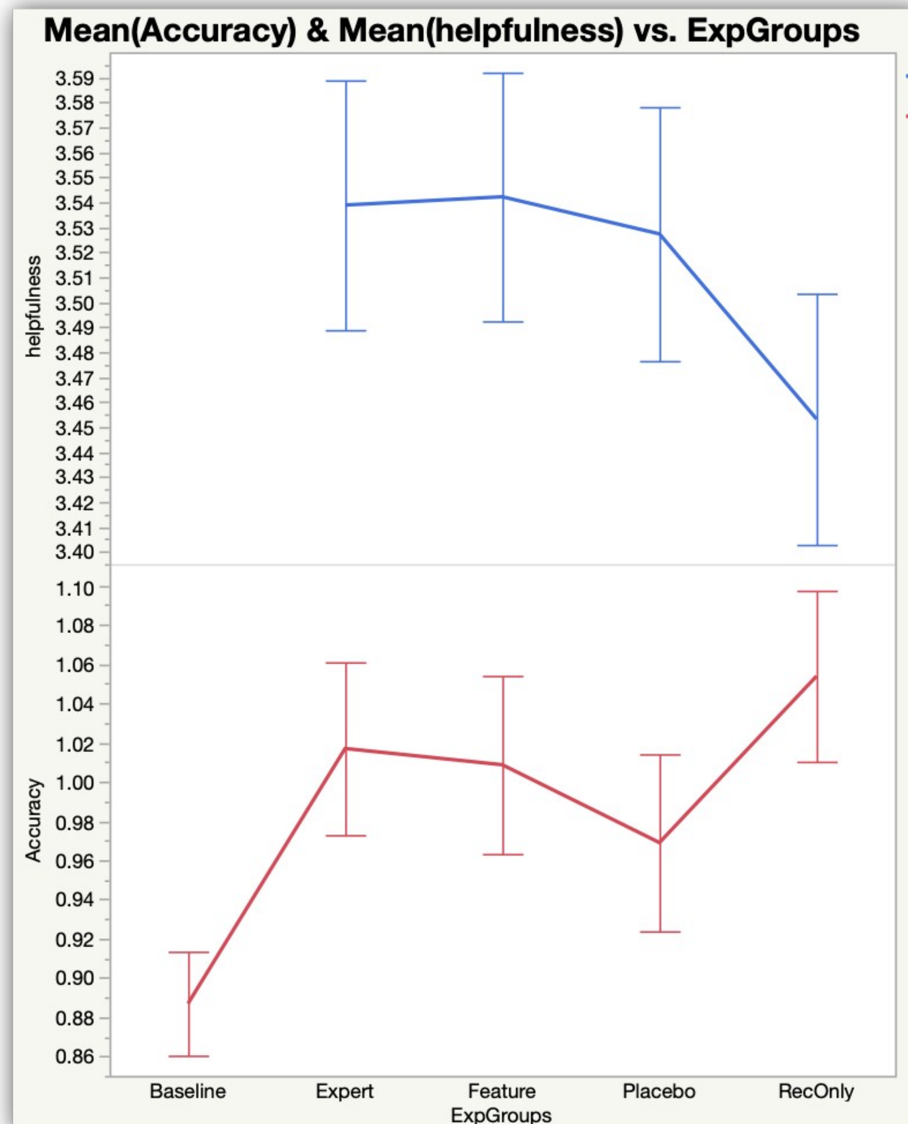


Within explanations: helpfulness, accuracy anti-correlated

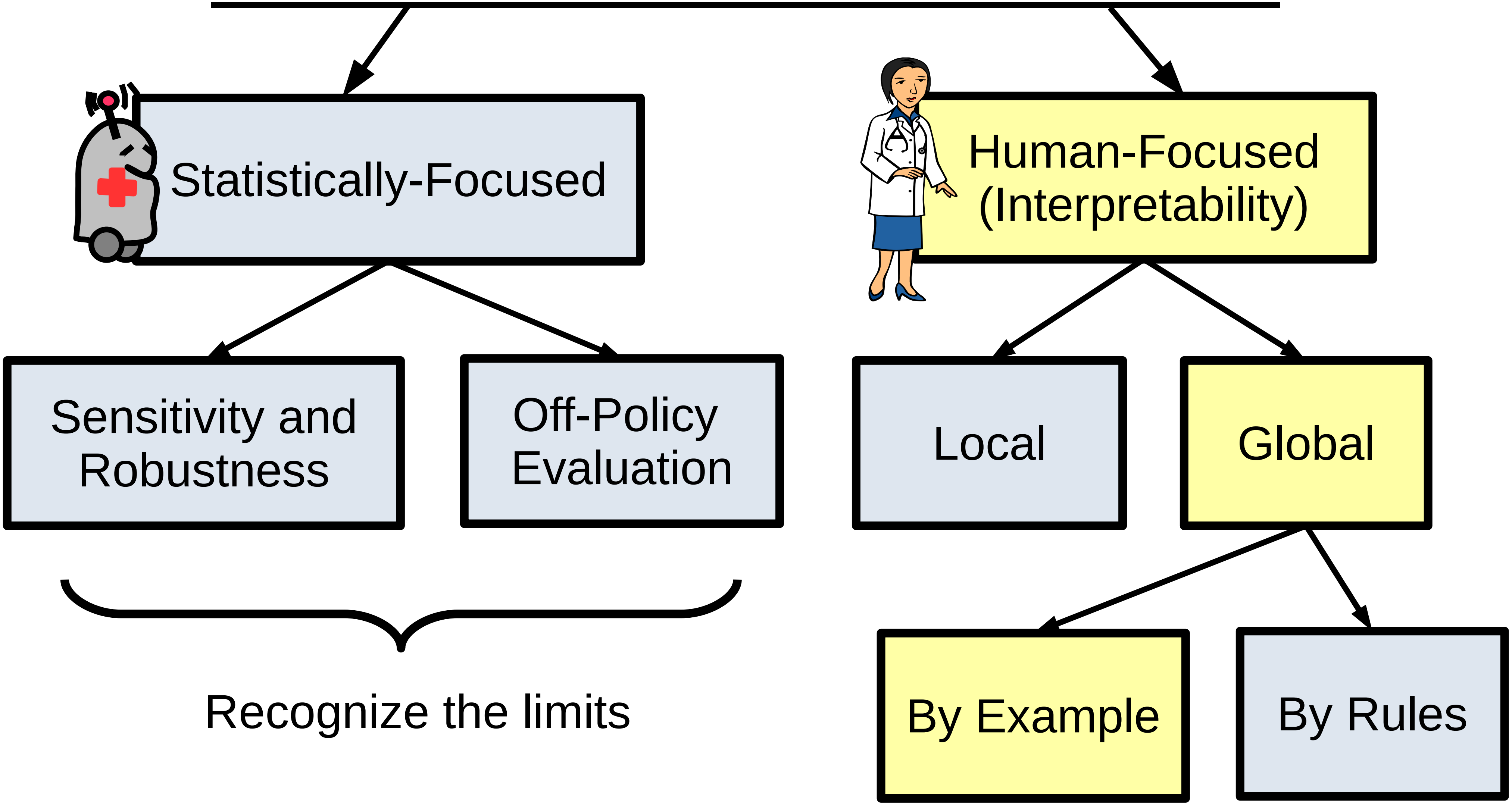
Recommendation incorrect



Recommendation correct



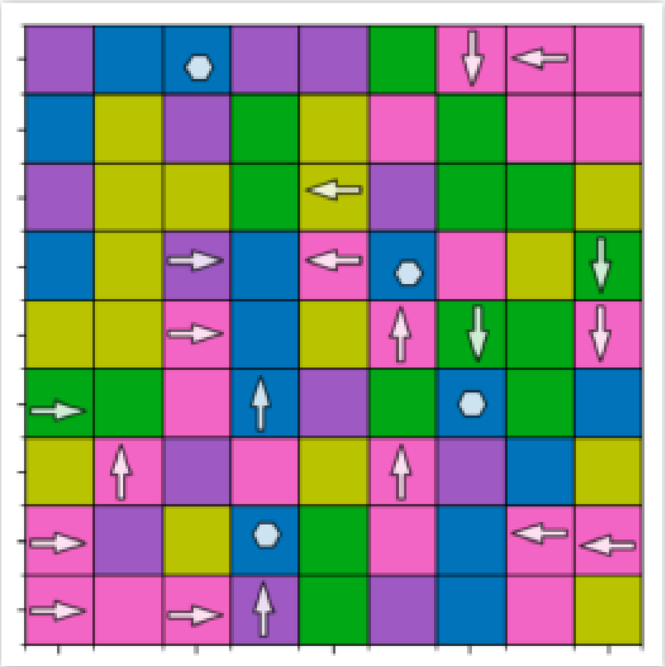
Batch Validation Roadmap



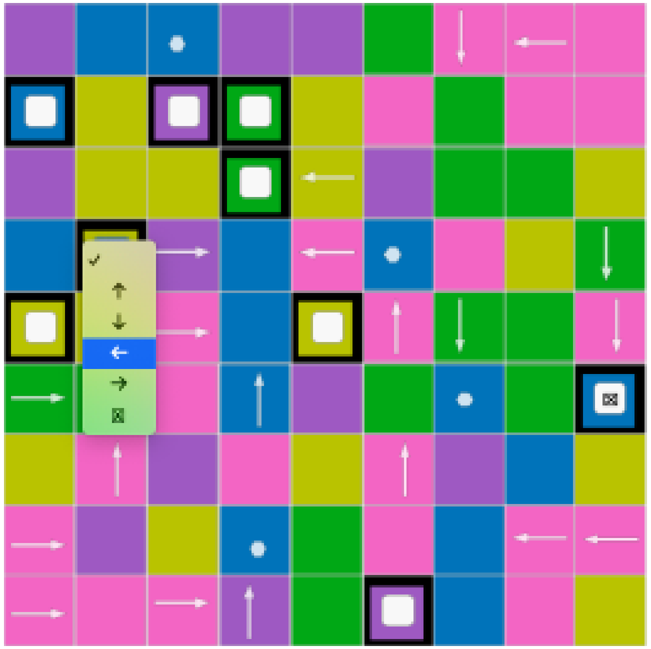
Can We Understand How People Process Examples?

Example: List some gridworld actions

Given:



Test: What happens here?

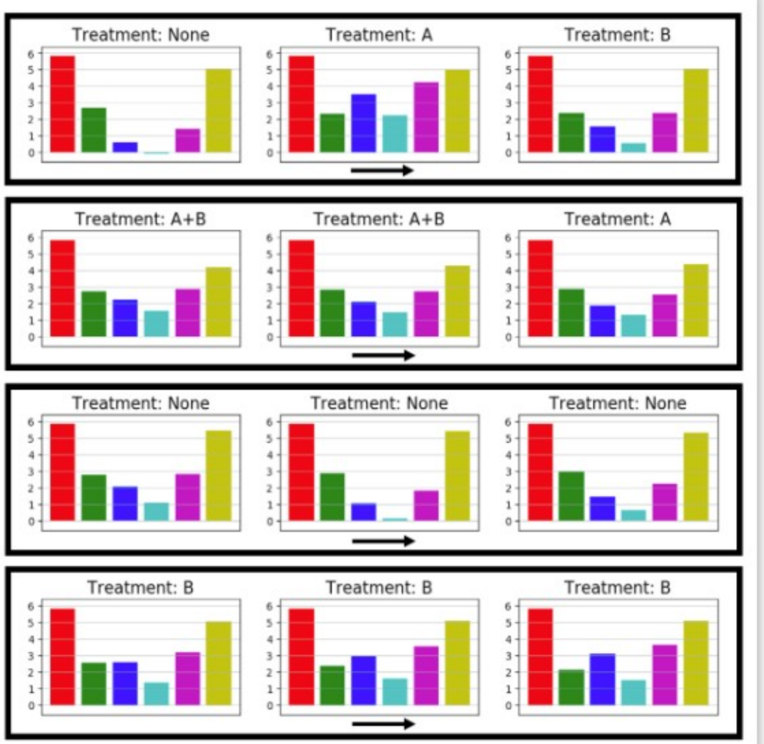


(Lage et al, IJCAI 2019)

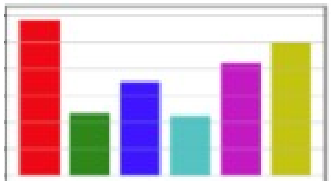
Can We Understand How People Process Examples?

Example: List some HIV actions

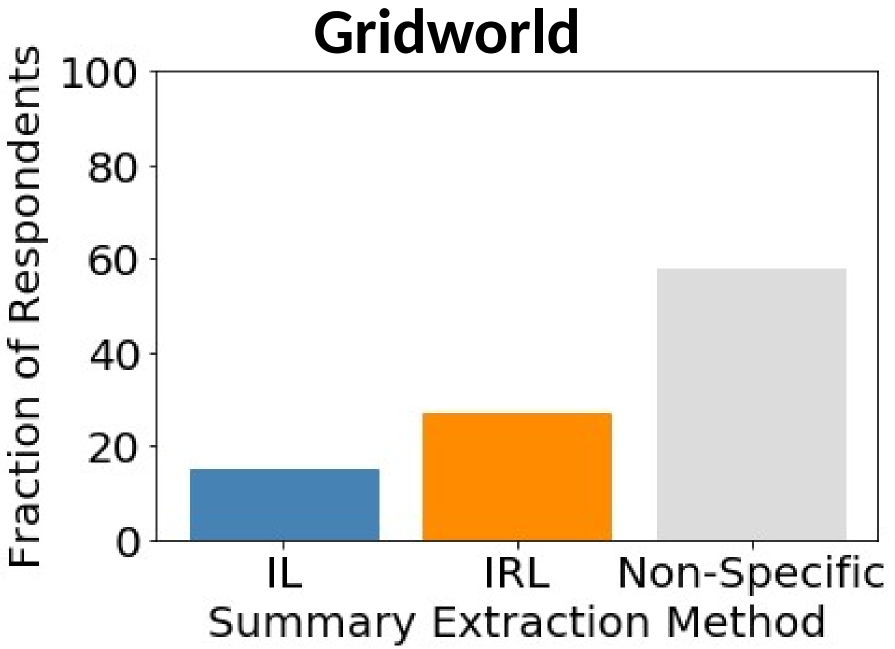
Given:

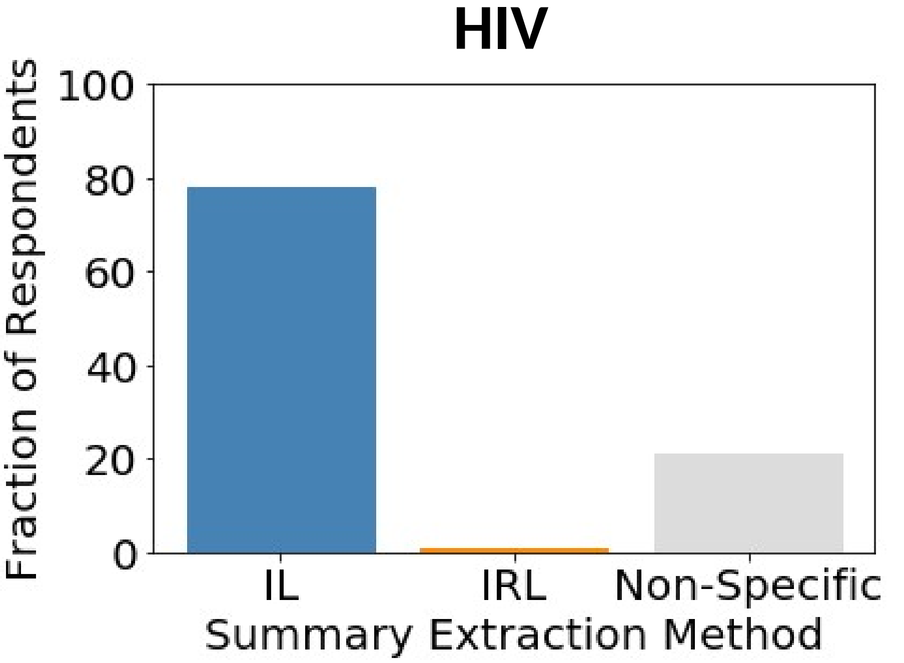


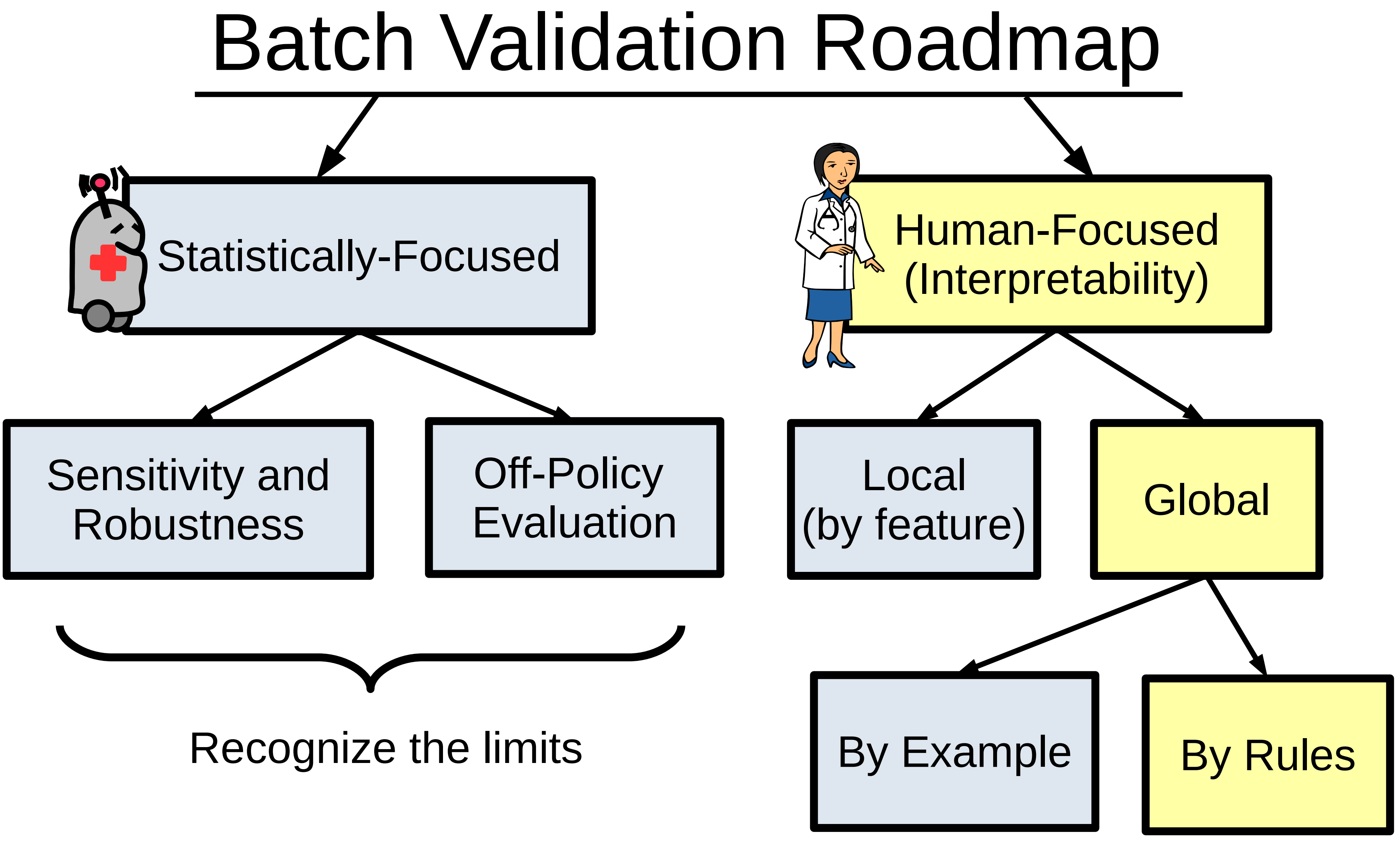
Test: What happens here?



Finding: Humans use different methods in different scenarios



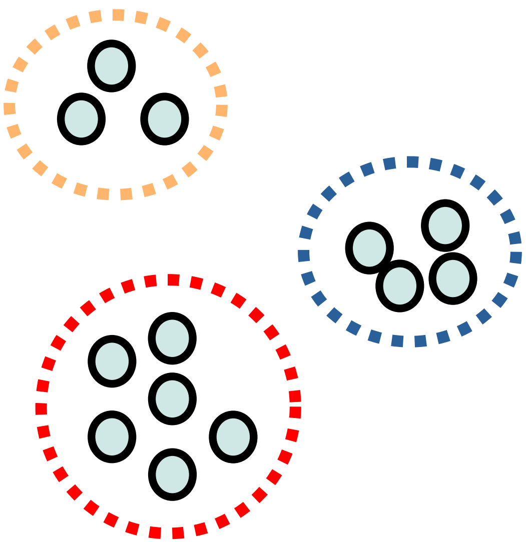




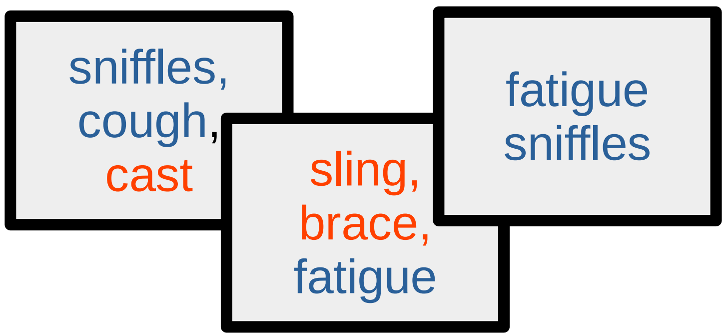
Batch Validation Roadmap
Statistically-Focused
Human-Focused (Interpretability)
Sensitivity and Robustness
Off-Policy Evaluation
Local (by feature)
Global
By Example
By Rules
Recognize the limits

Small, Interpretable Models

- Start: generative model (diseases create data).



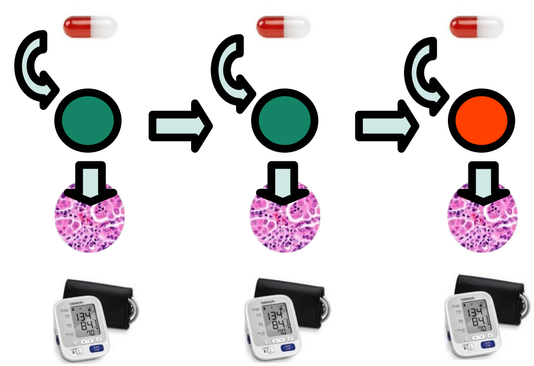
Mixture Model



Explained by topics:

Injury:	Cold:
cast	cough
sling	sniffles
brace	fatigue

Topic Model

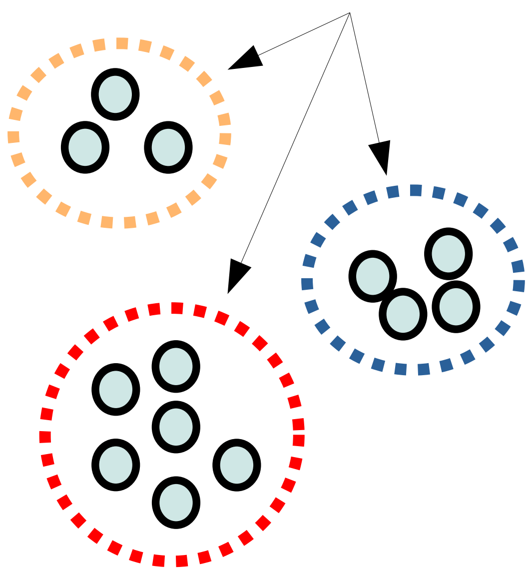


Partially-Observable
Markov Decision Process

Small, Interpretable Models

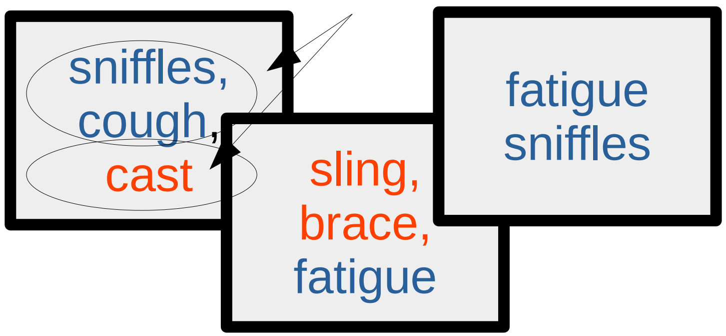
- Start: generative model (diseases create data).
- Train to be good at predictions.

Patient cluster predicts outcome



Mixture Model

Patients topics predict outcome

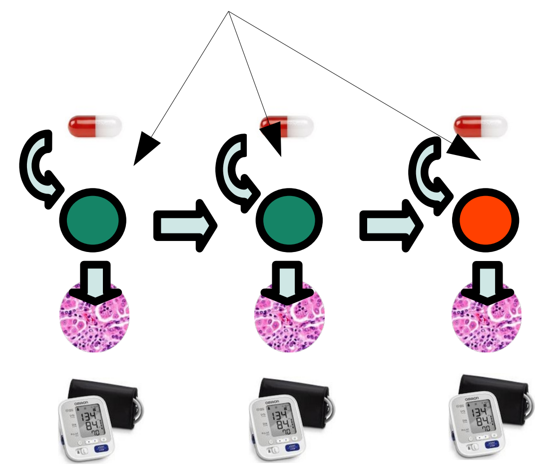


Explained by topics:

Injury:	Cold:
cast	cough
sling	sniffles
brace	fatigue

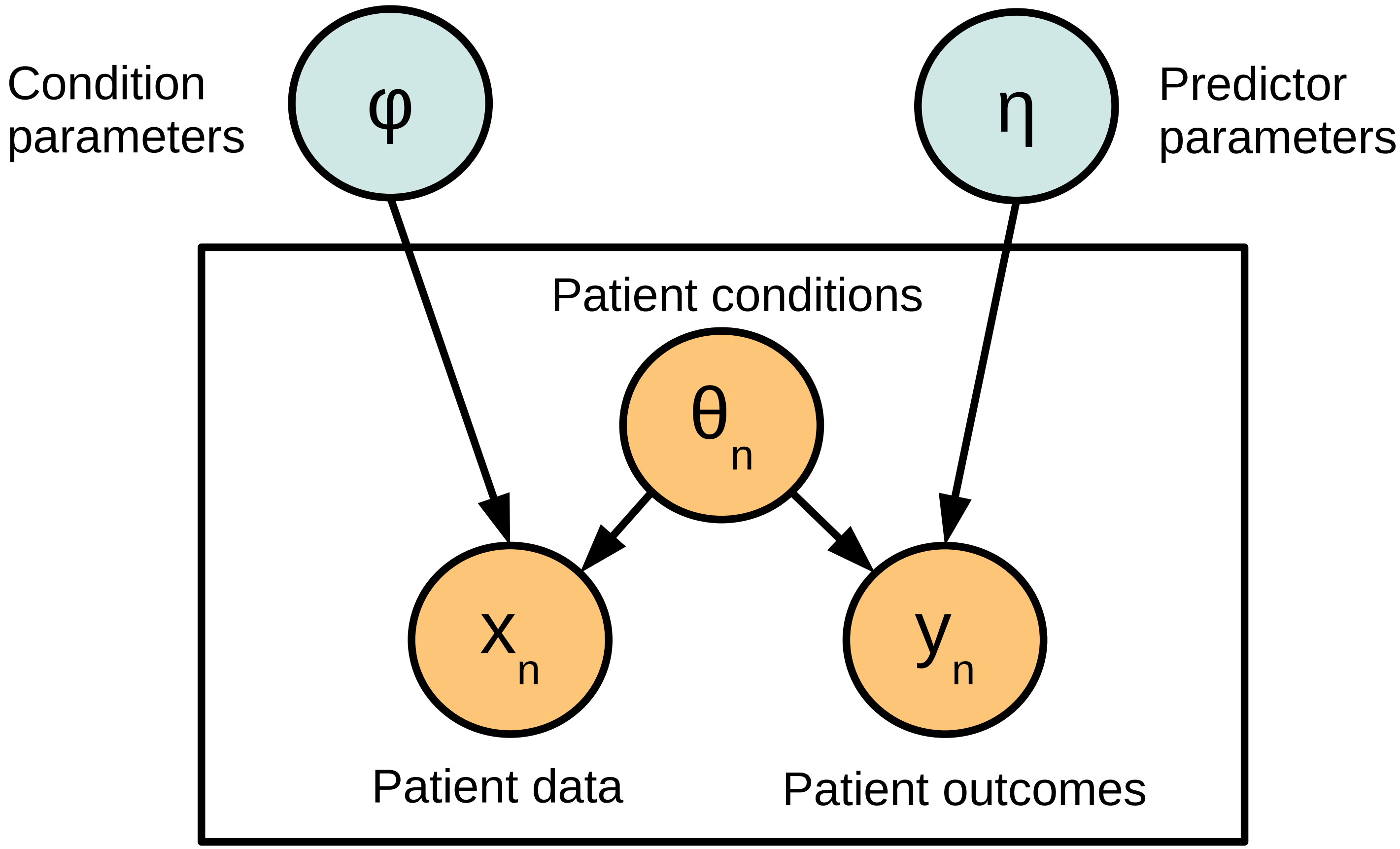
Topic Model

Patients state predicts progression

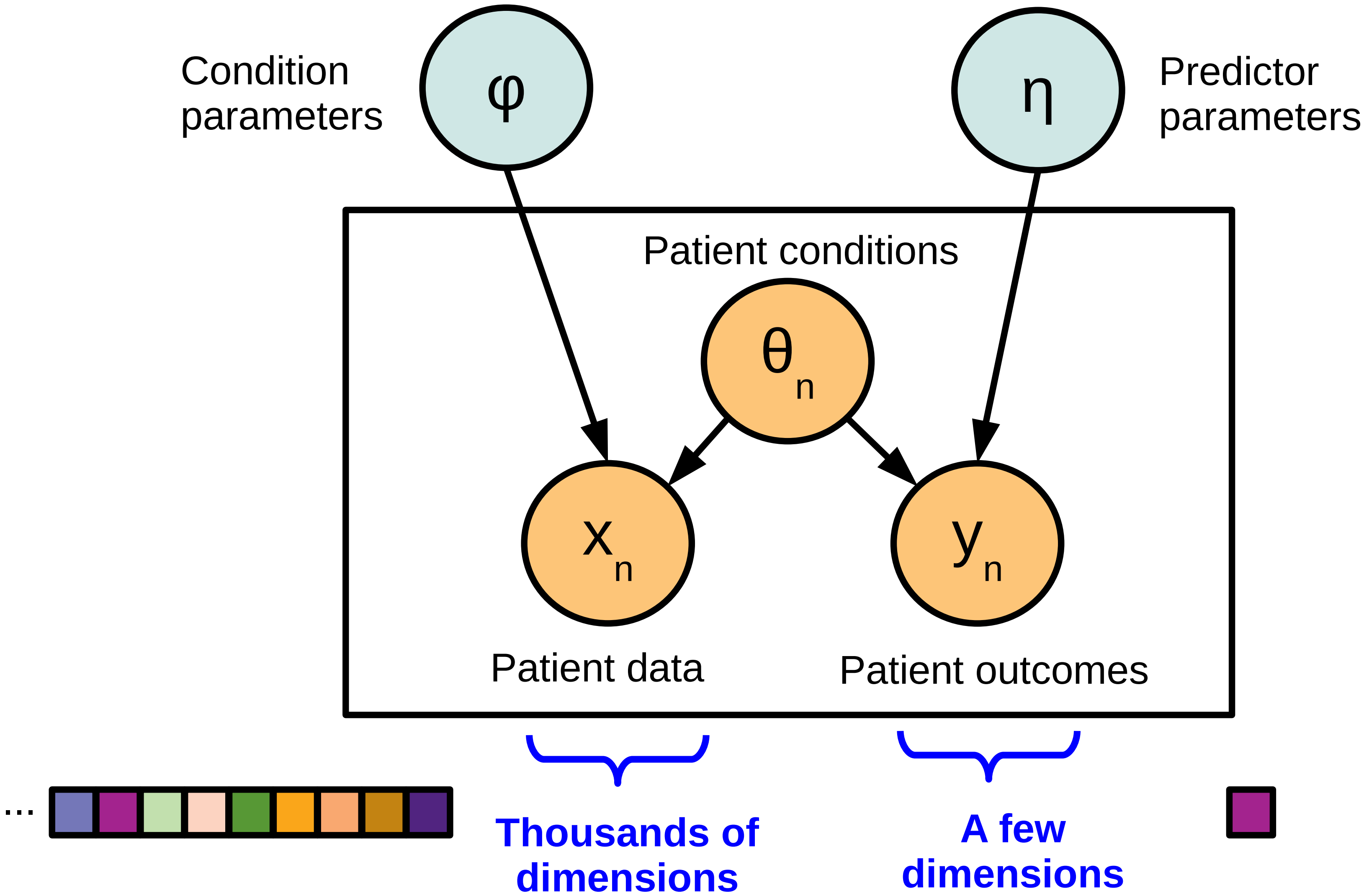


Partially-Observable Markov Decision Process

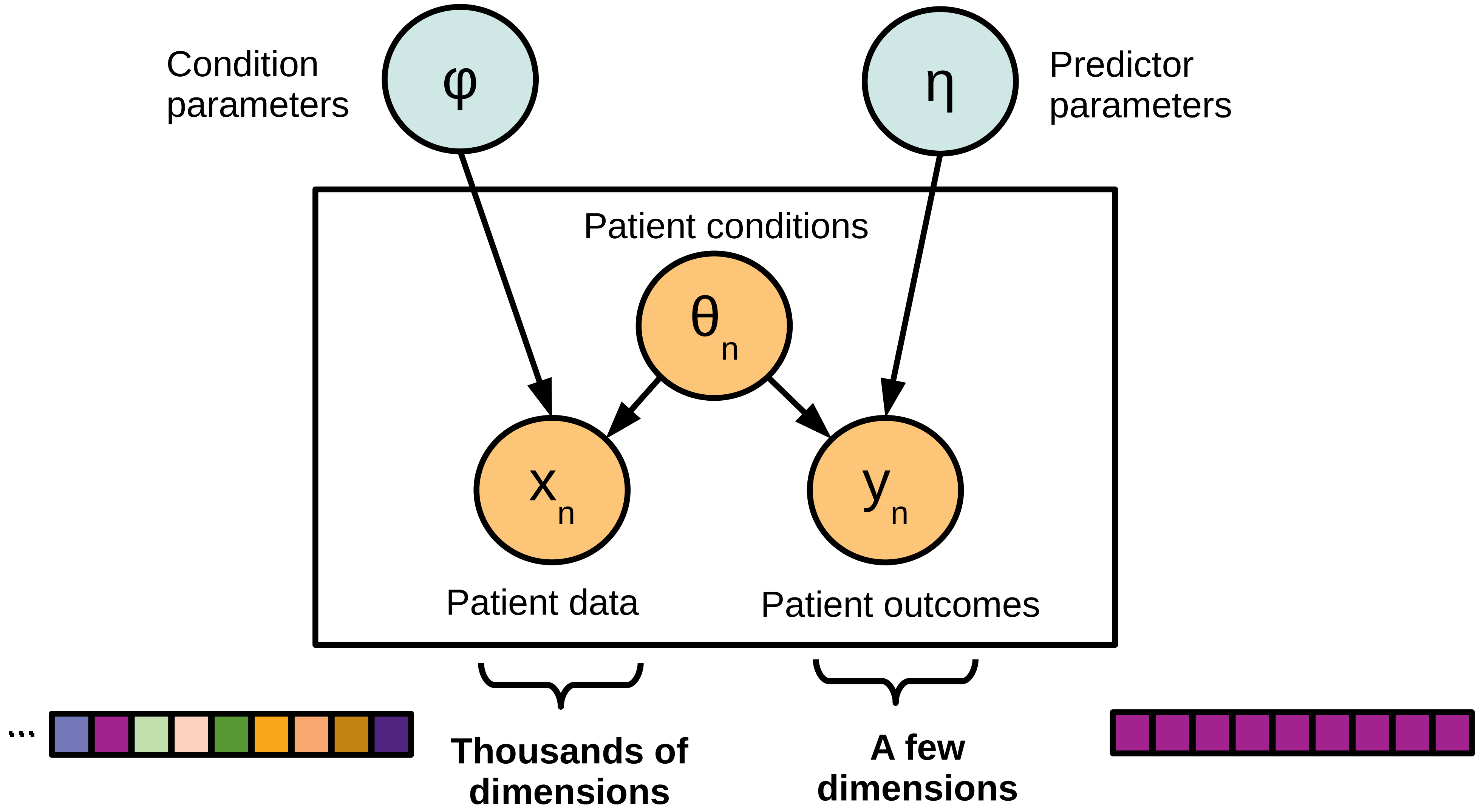
Formalizing this notion



Issue: Dimensionality of data, output



Issue: Dimensionality of data, output

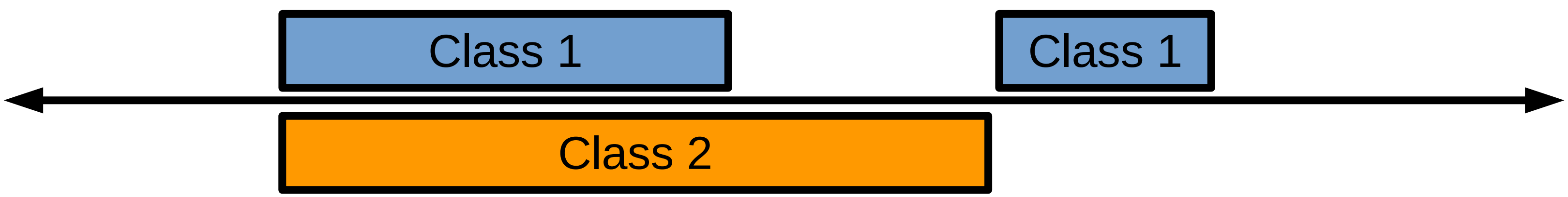


Previous Work Claim: Label replication will give good performance

Insight: Label replication isn't sufficient!

Replicating y does not capture the fact that we care about p(y|x) but not p(x|y).

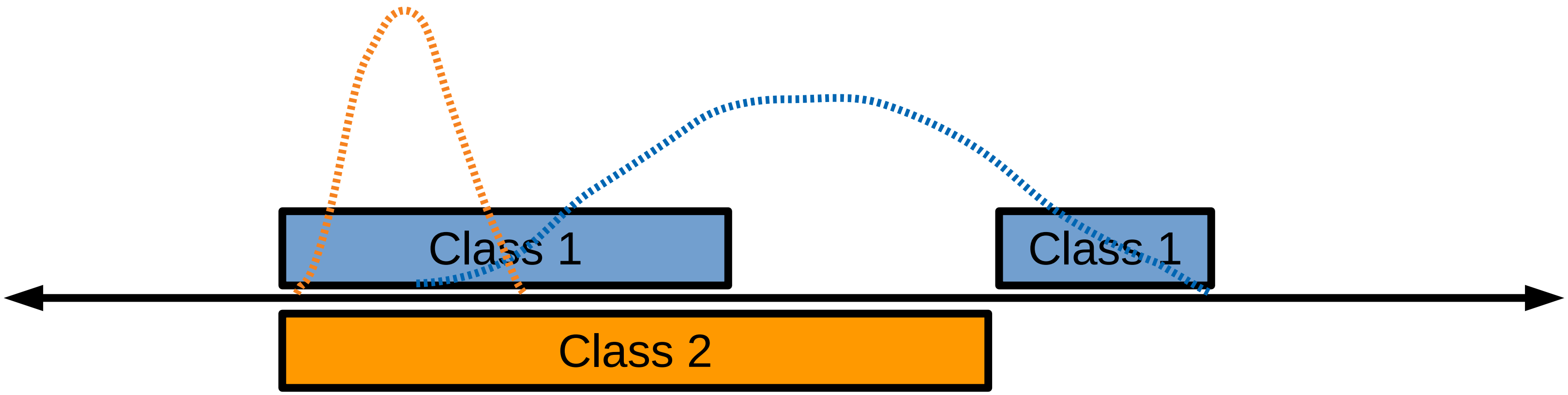
Thought experiment: Let's fit a discriminative mixture of two Gaussians to the following:



Insight: Label replication isn’t sufficient!

Replicating y does not capture the fact that we care about p(y|x) but not p(x|y).

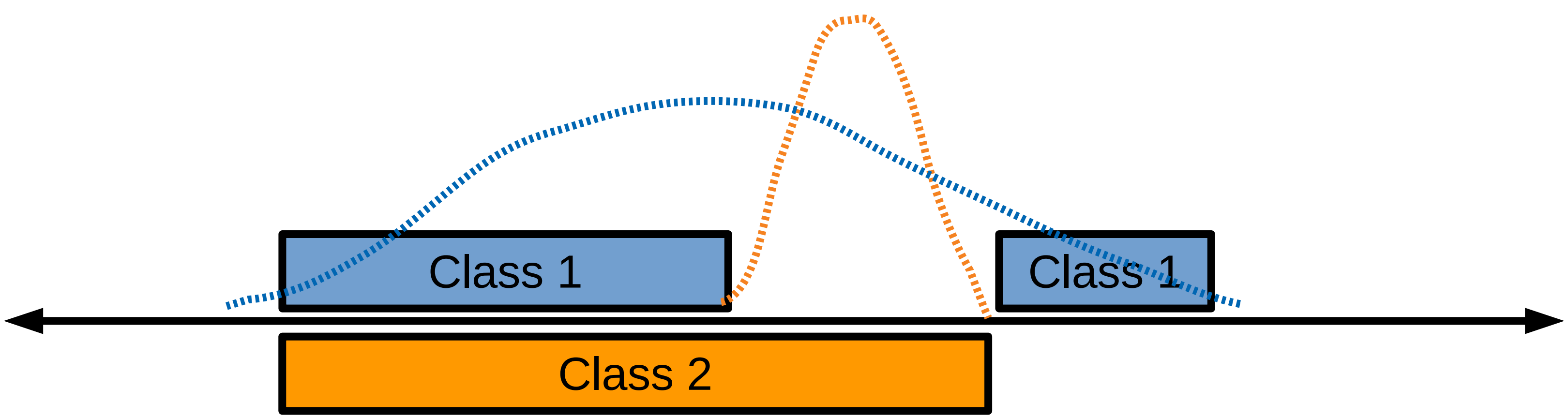
Thought experiment: Let’s fit a discriminative mixture of two Gaussians to the following:



Insight: Label replication isn't sufficient!

Replicating y does not capture the fact that we care about $p(y|x)$ but not $p(x|y)$.

Thought experiment: Let's fit a discriminative mixture of two Gaussians to the following:

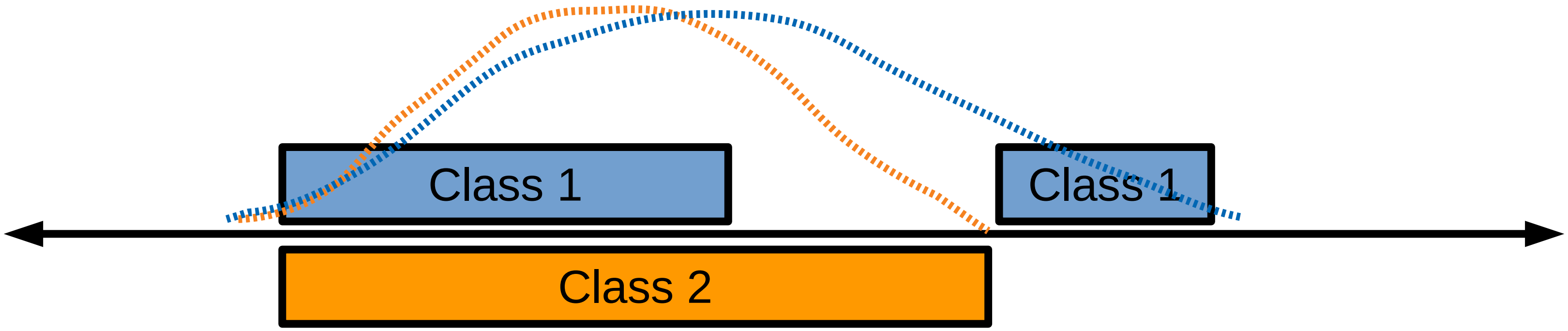


A reasonable solution…

Insight: Label replication isn’t sufficient!

Replicating y does not capture the fact that we care about p(y|x) but not p(x|y).

Thought experiment: Let’s fit a discriminative mixture of two Gaussians to the following:



Note: fitting data distributions isn’t best!

A little math

Joint likelihood:

$$p(x, y|\phi, \eta) = \prod_n \int_{\theta_n} p(x_n|\phi, \theta_n)\, p(y_n|\eta, \theta_n)\, p(\theta_n)$$

Joint likelihood with replication:

$$p(x, y|\phi, \eta) = \prod_n \int_{\theta_n} p(x_n|\phi, \theta_n)\, p(y_n|\eta, \theta_n)^R\, p(\theta_n)$$

A little math

Joint likelihood:

$$p(x, y|\phi, \eta) = \prod_n \int_{\theta_n} p(x_n|\phi, \theta_n)\, p(y_n|\eta, \theta_n)\, p(\theta_n)$$

Joint likelihood with replication:

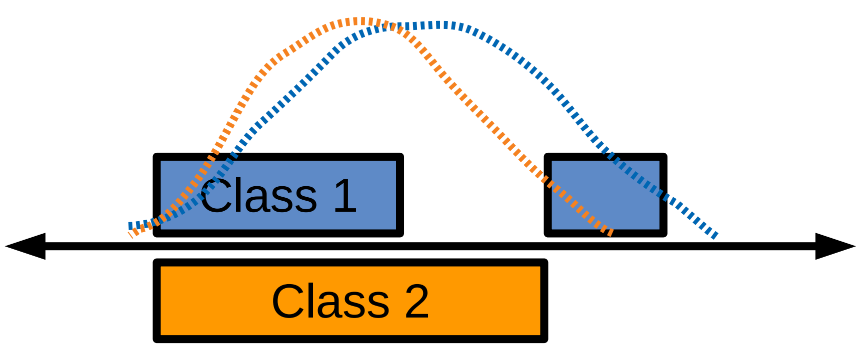
$$p(x, y|\phi, \eta) = \prod_n \int_{\theta_n} p(x_n|\phi, \theta_n)\, p(y_n|\eta, \theta_n)^R\, p(\theta_n)$$

When R is large, large pressure for θ_n to be a perfect predictor of y_n… but no pressure for x_n to be a predictor of y_n

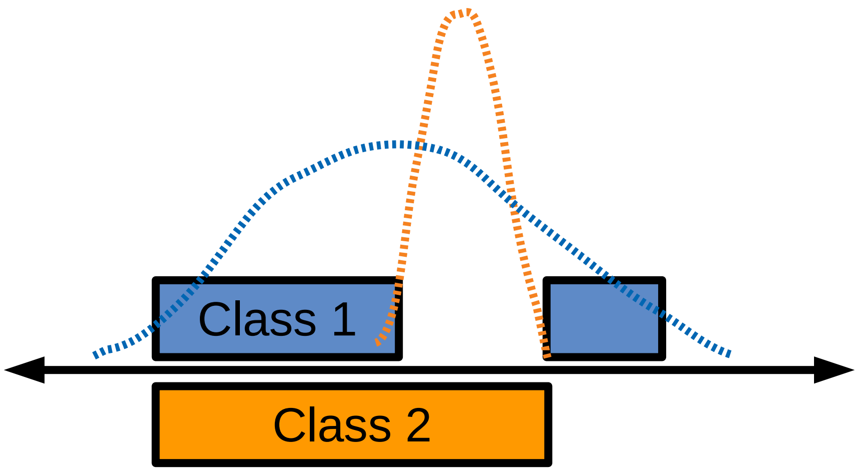
Insight: Label replication isn't sufficient!

Replicating y does not capture the fact that we care about p(y|x) but not p(x|y).

Thought experiment: Let's fit a discriminative mixture of two Gaussians to the following:



Replication solution

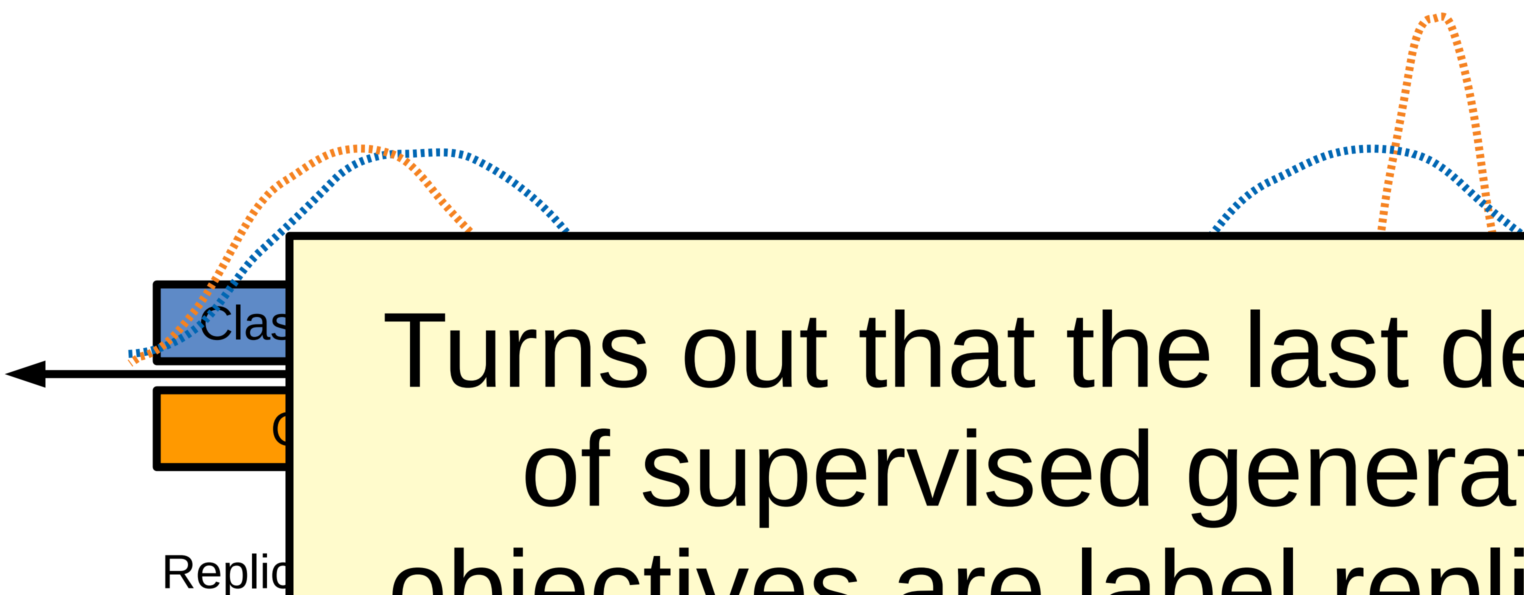


Better Solution

Insight: Label replication isn't sufficient!

Replicating y does not capture the fact that we care about p(y|x) but not p(x|y).

Thought experiment: Let's fit a discriminative mixture of two Gaussians to the following:



Turns out that the last decade+ of supervised generative objectives are label replication!

Our Solution: Task-Constrained Objective

$$\left.min_{\phi,\eta} - \sum_n \log p(x_n|\phi)\right\}$$ Explain the data the best you can

Subject to

$$\left.-\sum_n \log p(y_n|x_n,\phi,\eta) < L\right\}$$ While making good predictions

Different than label replication!

Label replication:

$$min_{\phi,\eta} - \sum_n \log \int_{\theta_n} p(x_n|\phi,\theta_n)\, p(y_n|\eta,\theta_n)^R\, p(\theta_n)$$

Prediction-Constrained Objective

$$min_{\phi,\eta} - \sum_n \log \int_{\theta_n} p(x_n|\theta_n,\phi)\, p(\theta_n)$$
$$+\lambda \log \int_{\theta_n} p(y_n|\theta_n,\eta)\, p(\theta_n|x_n)$$

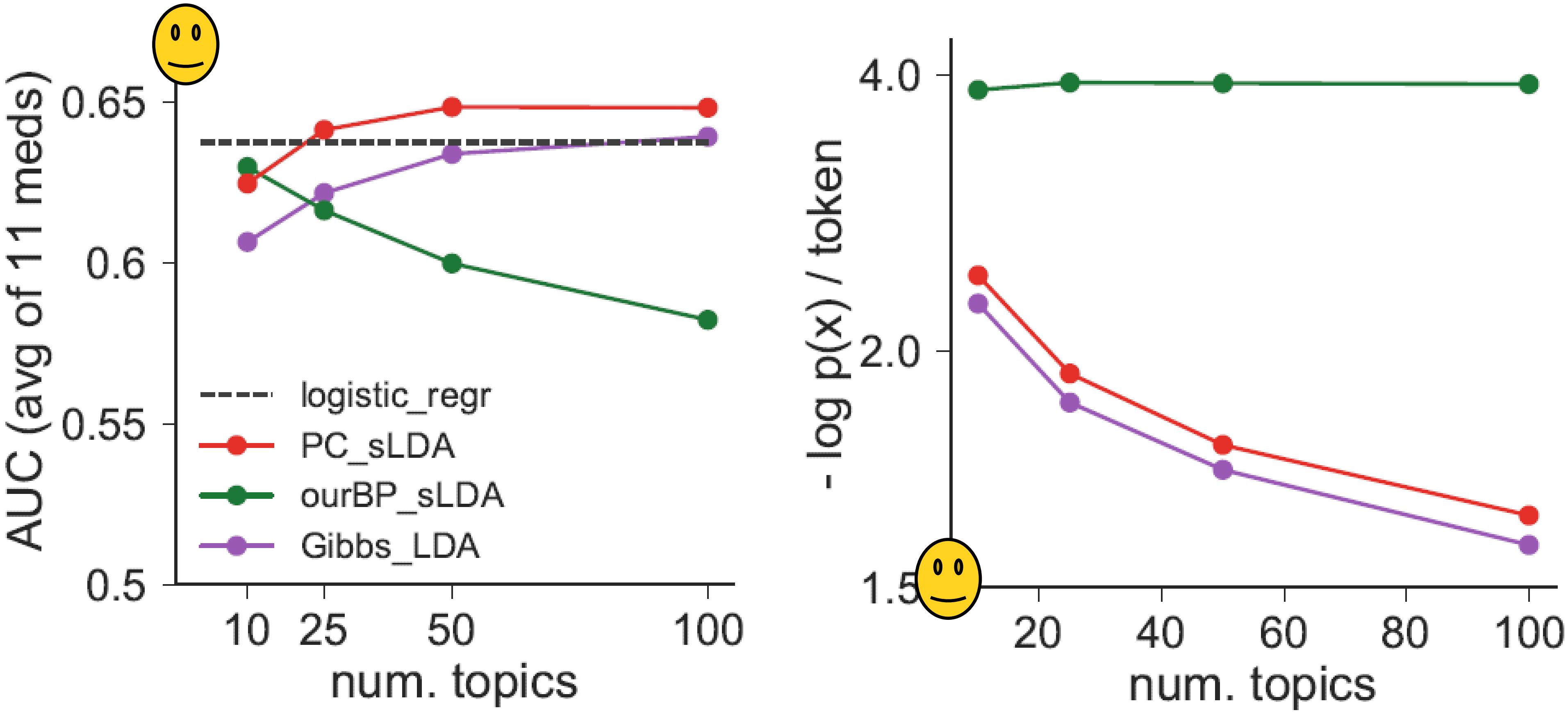
Our task-constrained objective replicates the target **task**, not the target.

Application: Antidepressant Selection with Small Topic Models

- Inputs: 7,291 common health record codes
- Actions: 10 common antidepressants
- Goal: Identify drugs that will work (stable over 90 days) for each person
- Approach: Reduce dimensionality with topic models, use those topics to recommend actions.

(Hughes et al., AISTATS 2018)

Application: Antidepressant Selection with Small Topic Models



Application: Antidepressant Selection with Small Topic Models

Decision only

BPsLDA +7.7

0.60 nortriptyline
0.27 nonspecific abnormal findings
0.21 other specified local infection
0.20 embrionic cyst of the fallopian tube
0.18 application of the intervertebrea...
0.16 other malignant neoplasm...
0.15 amoxicilllin/clarithromycin
0.15 need for prophylactic vaccine

Data only

Gibbs -0.6

1.0000 bipolar, depressive
0.9999 bipolar, unspecified
0.9999 schizo-affective schizophrenia
0.9999 bipolar, mixed
0.9998 electroconvulsive therapy
0.9998 anesthesia for ECT
0.9997 residual schizophrenia
0.9996 other electroshock therapy

Both

PCLDA +3.8

0.99 migraine, unspecified, without...
0.99 other malaise and fatigue
0.99 common migraine...
0.99 sumatriptan
0.99 asa/butalbital/caffeine
0.99 zolmitriptan
0.99 migraine, unspecified, with..,
0.99 classical migraine, without...
0.99 classical migraine, with...

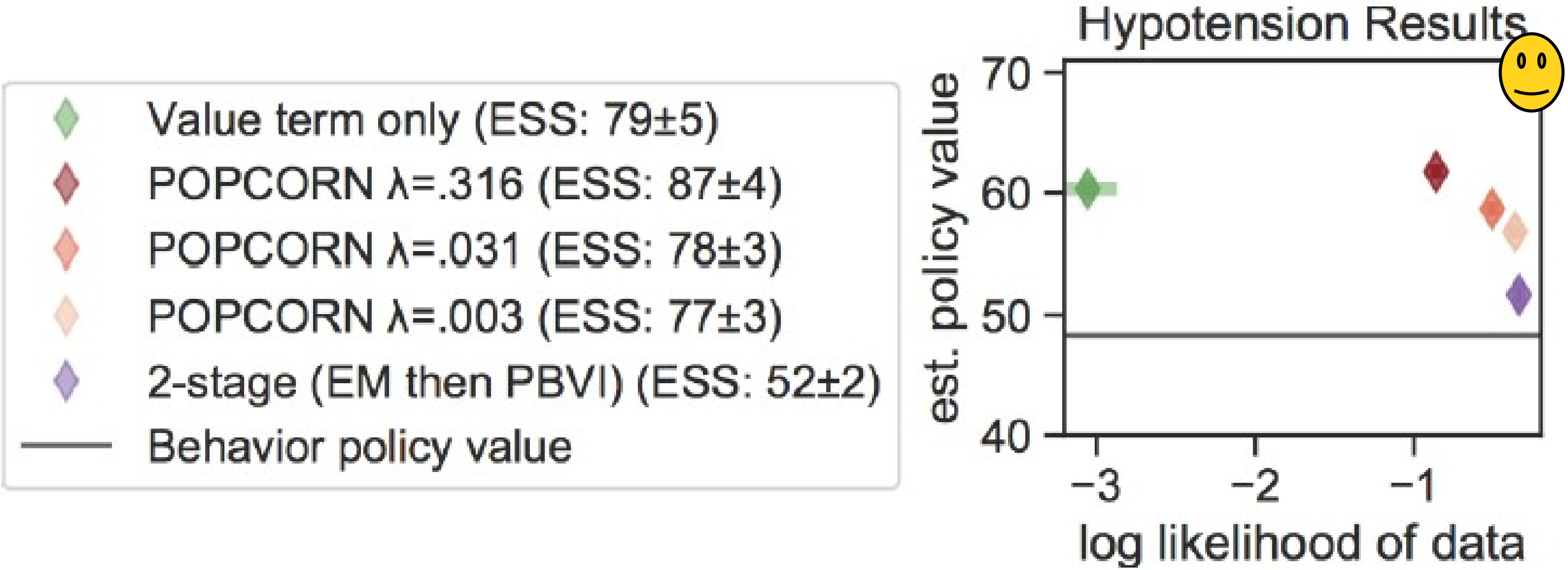
Application: Hypotension Management with Small POMDPs

- Inputs: 9 vitals/labs over 72 hours in ICU for 10K stays
- Actions: discretized fluid, vasopressor administration
- Goal: keep blood pressure in range
- Approach: Learn a small, discrete POMDP to recommend actions.

Objective = LogLikelihood(data w.r.t. m)
+ λOffPolicyValueEstimate(π* w.r.t. m)

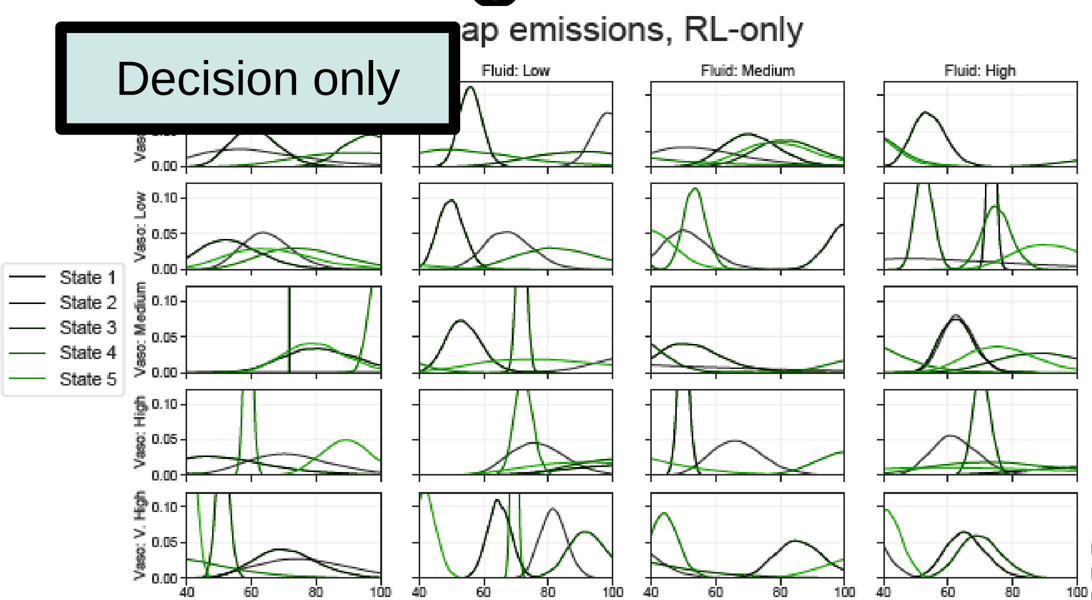
(Futoma et al., AISTATS 2020)

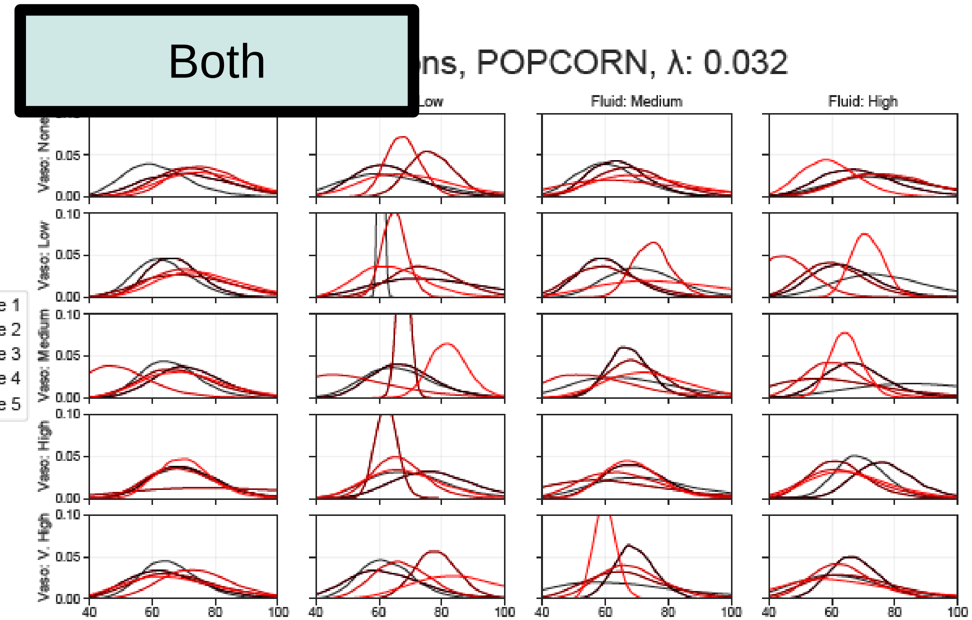
Application: Hypotension Management with Small POMDPs

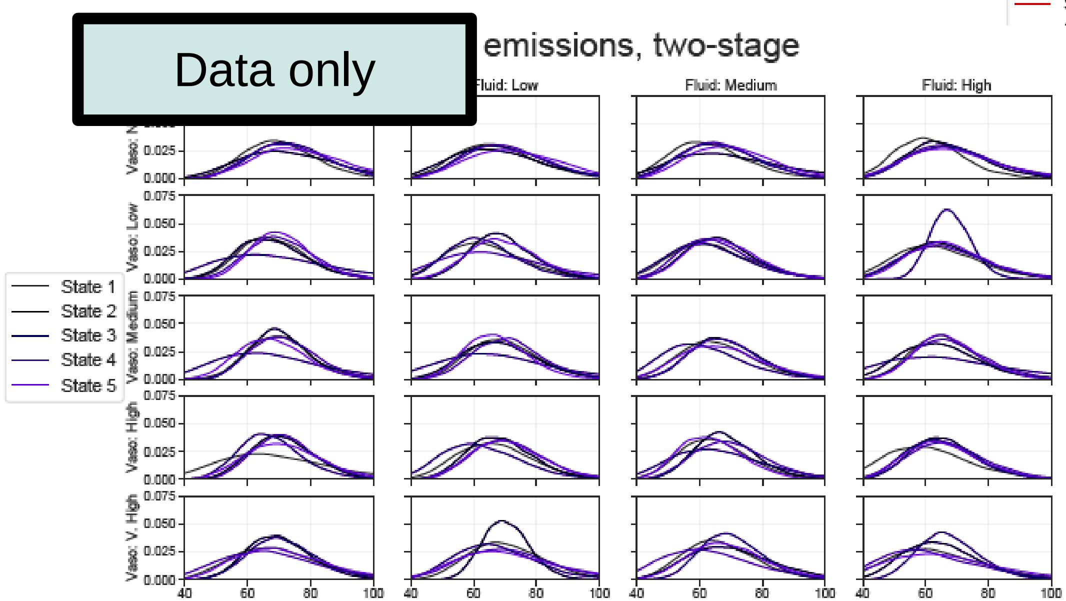


And only 5 discrete states! Example in the story had 128 continuous states.

Application: Hypotension Management with Small POMDPs



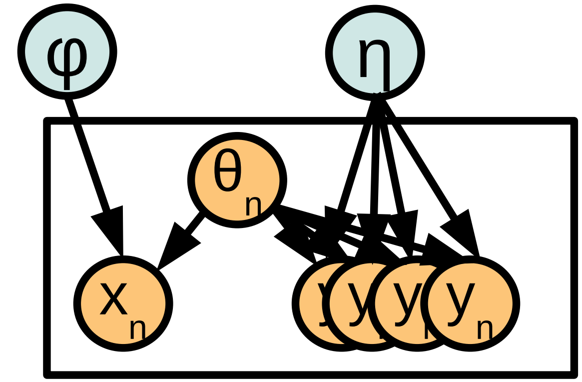




Efficient Inference Insight

Label replication:

$$min_{\phi,\eta} - \sum_n \log \int_{\theta_n} p(x_n|\phi,\theta_n)\, p(y_n|\eta,\theta_n)^R\, p(\theta_n)$$



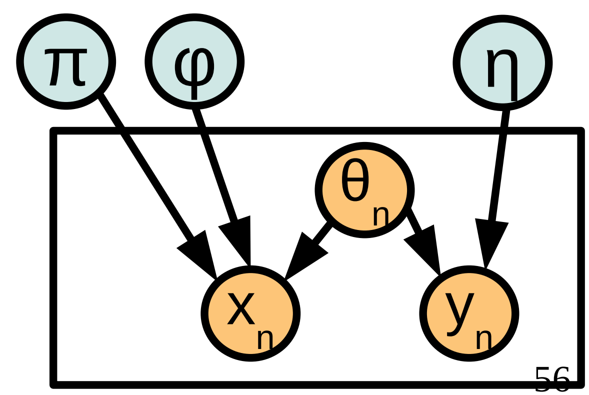
Prediction-Constrained Objective:

$$min_{\phi,\eta} - \sum_n \log p(x_n|\phi) + \lambda \log p(y_n|x_n,\phi,\eta)$$

No Model :(

Prediction-Focused Objective:

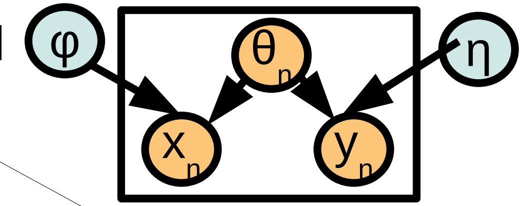
$$\begin{aligned} min_{\phi,\eta} - \sum_n & (1-p) \log p(x_n|\pi) \\ & + p \log p(x_n|\phi) \\ & + E[\log p(y_n|x_n,\phi,\eta)] \end{aligned}$$



Ren et al. AISTASTS 2020

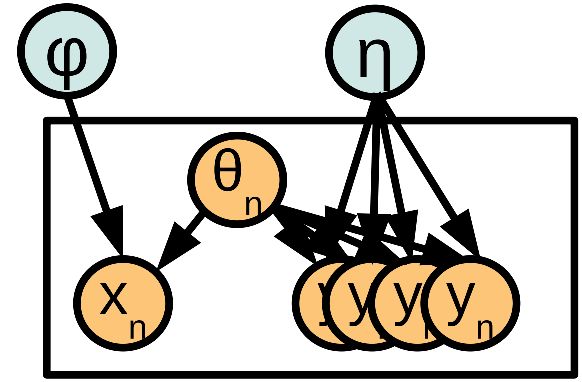
Efficient Inference Insight

Original



Label replication:

$$min_{\phi,\eta} - \sum_n \log \int_{\theta_n} p(x_n|\phi,\theta_n)\, p(y_n|\eta,\theta_n)^R\, p(\theta_n)$$



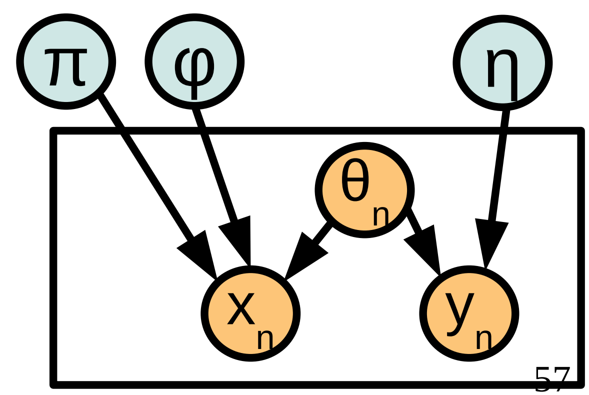
Prediction-Constrained Objective:

$$min_{\phi,\eta} - \sum_n \log p(x_n|\phi) + \lambda \log p(y_n|x_n,\phi,\eta)$$

No Model :(

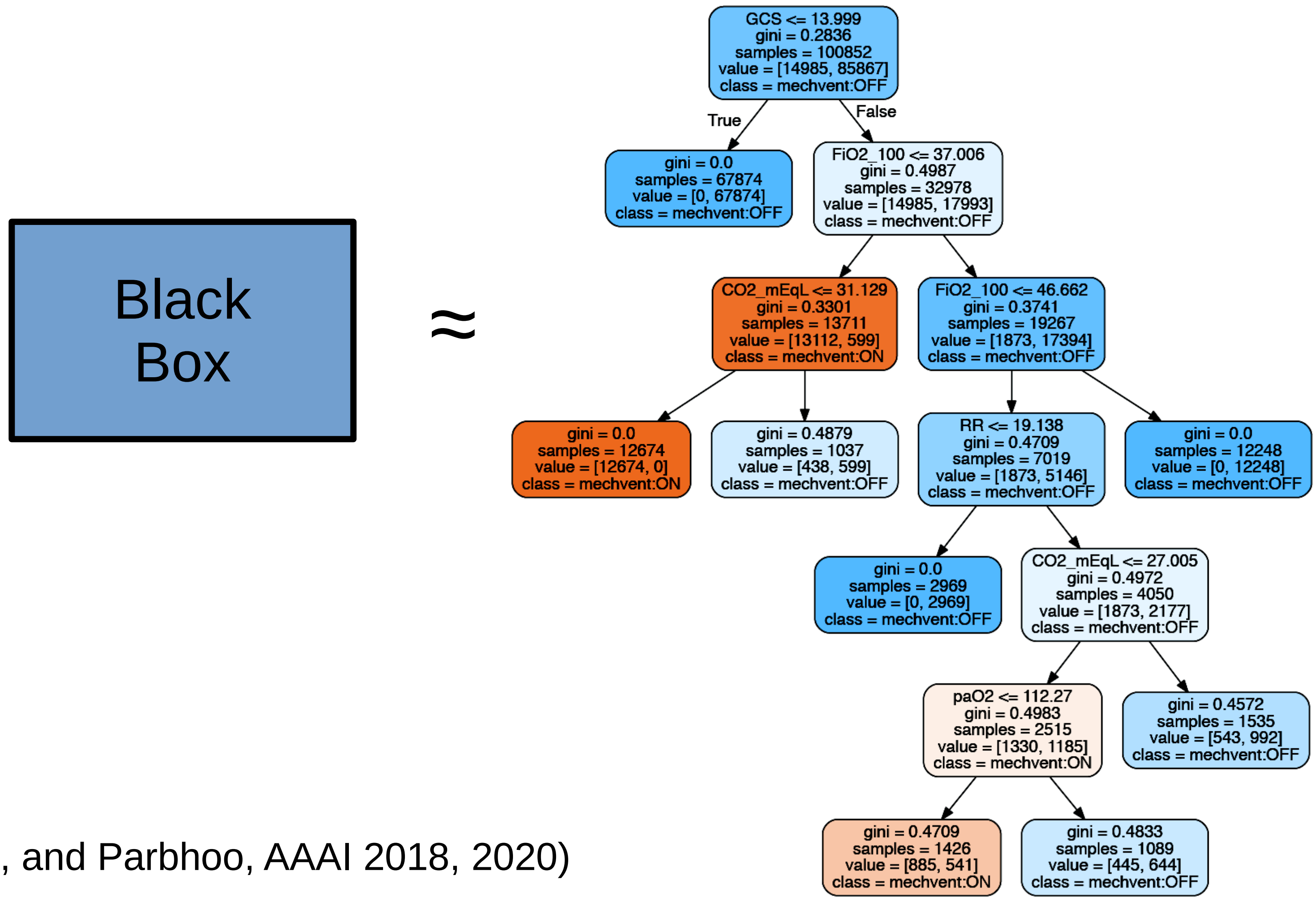
Prediction-Focused Objective:

$$\begin{aligned} min_{\phi,\eta} - \sum_n & (1-p) \log p(x_n|\pi) \\ & + p \log p(x_n|\phi) \\ & + E[\log p(y_n|x_n,\phi,\eta)] \end{aligned}$$



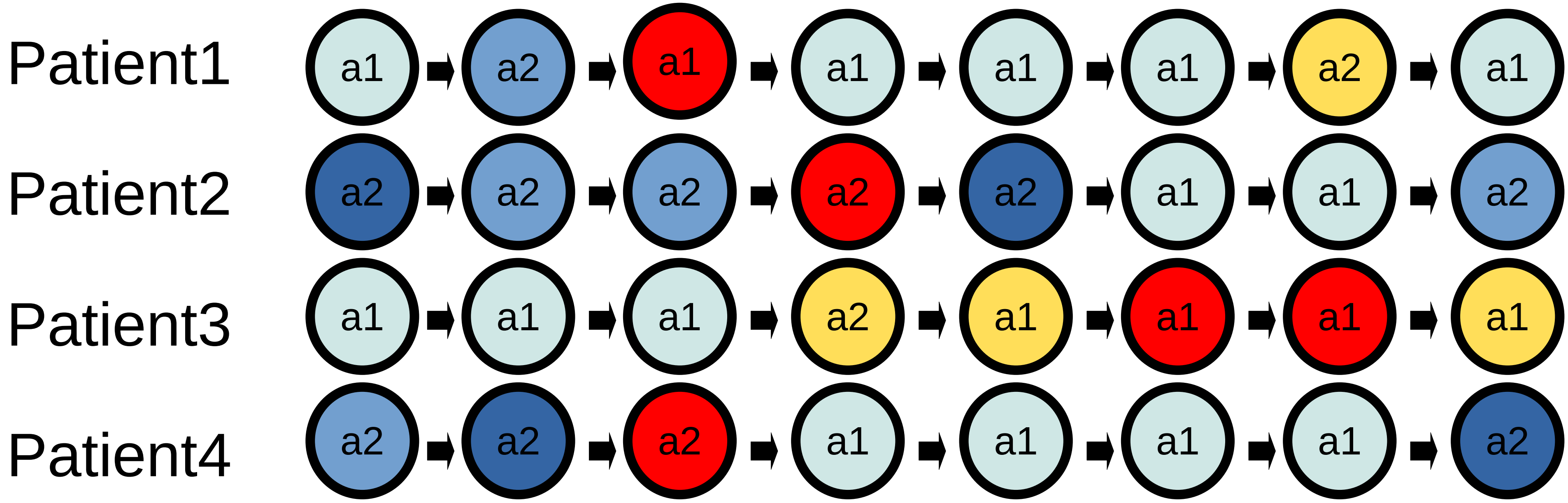
Ren et al. AISTASTS 2020

More AI to make small models: Models Close to Decision Trees

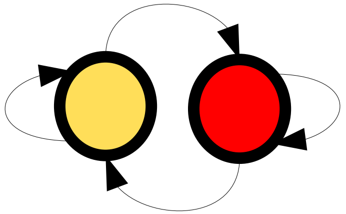


(Wu, Hughes, and Parbhoo, AAAI 2018, 2020)

More AI to make small models: Optimize only when Doctors Disagree



Just two states to optimize; we can build a tiny 2-state MDP!



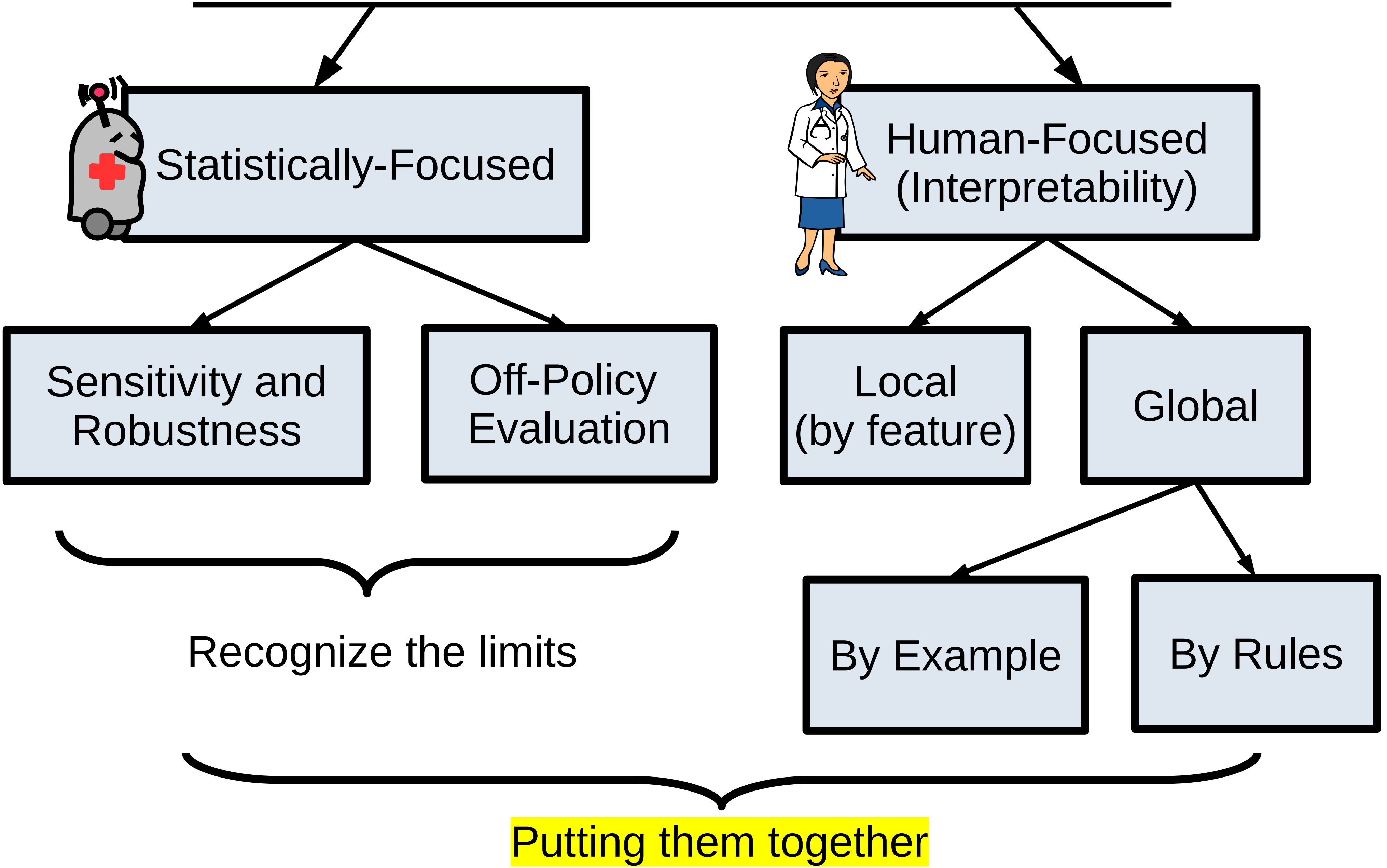
Zhang et al. ML4H WS 2020

More AI to make small models: Optimize only when Doctors Disagree

Back to Hypotension: Policy Summaries

Zhang et al. ML4H WS 2020

Batch Validation Roadmap



Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

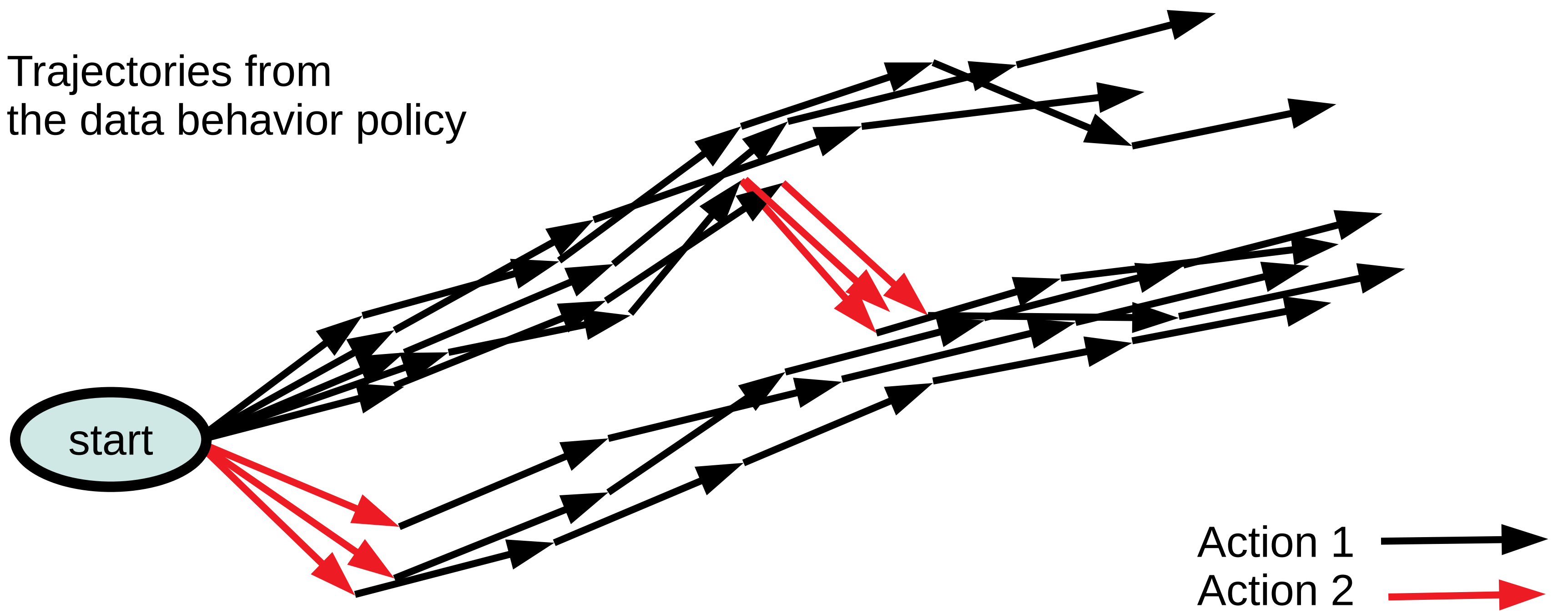
Core idea: Expose sensitive points to humans to validate.

(Gottesman et al, ICML 2020)

Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

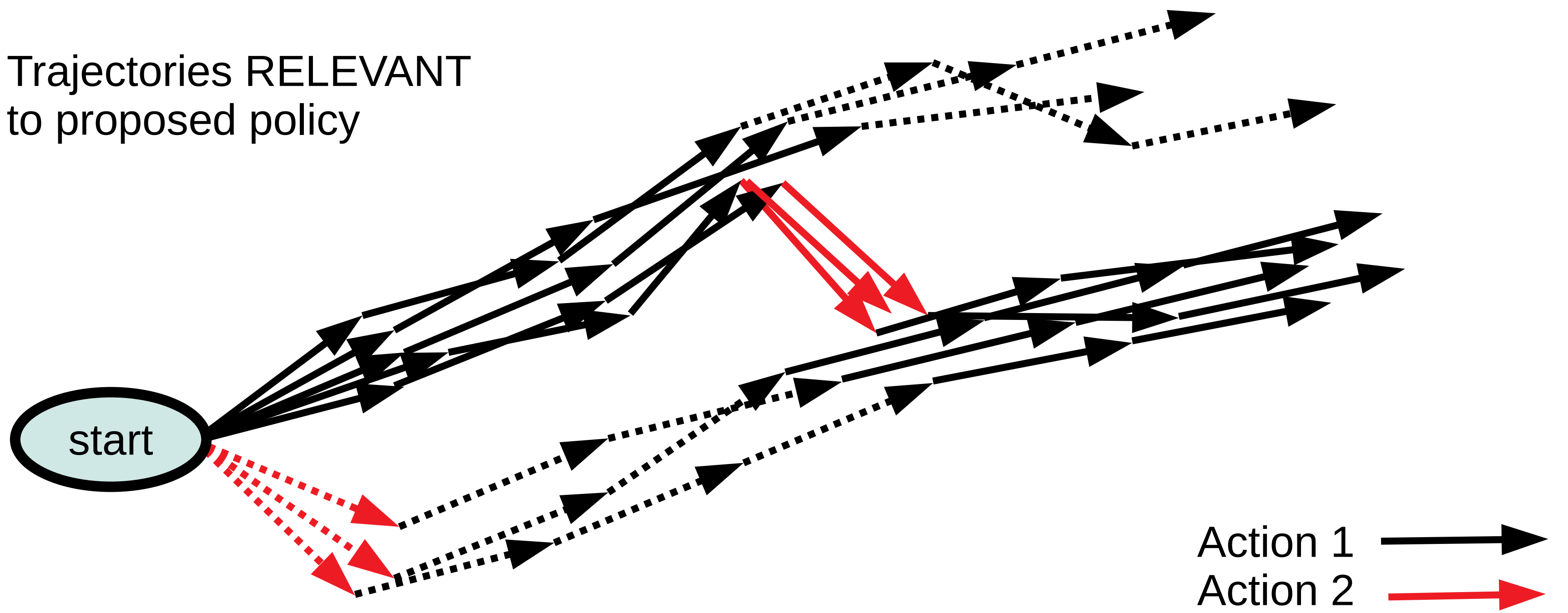
Core idea: Expose sensitive points to humans to validate.



Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

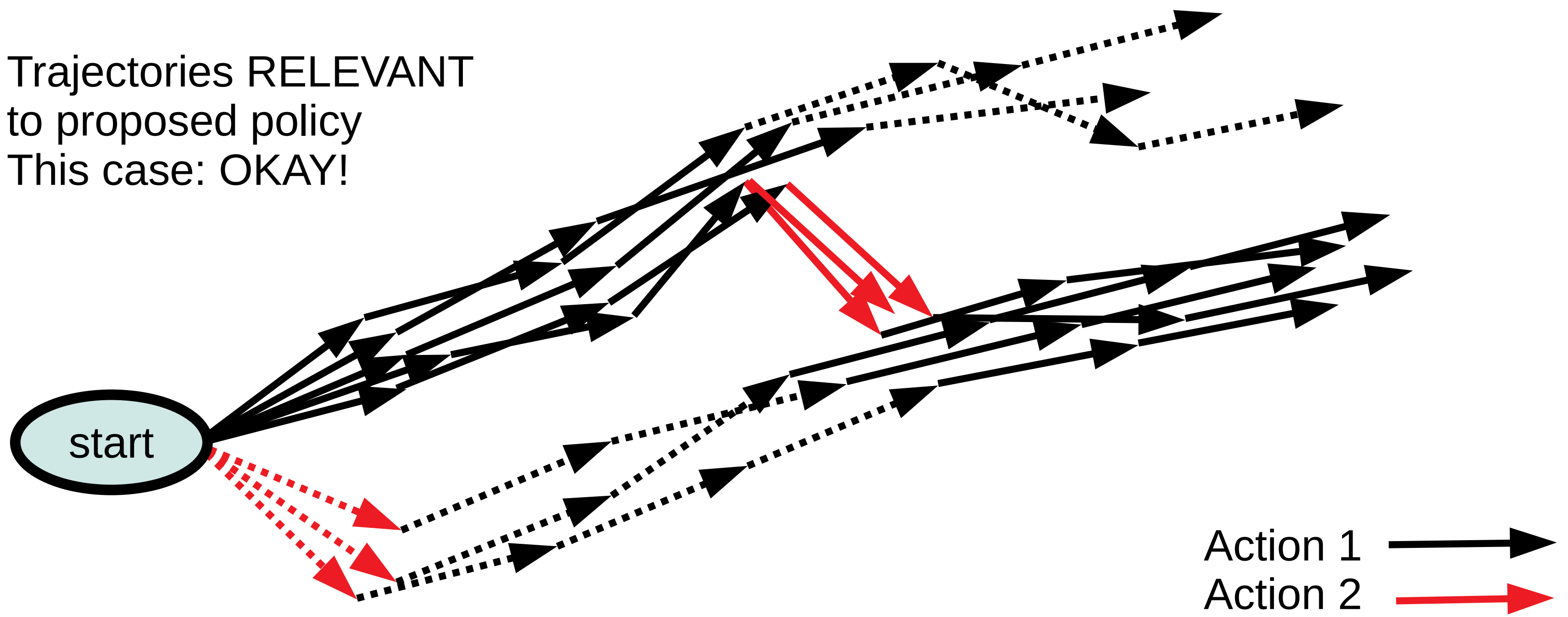
Core idea: Expose sensitive points to humans to validate.



Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

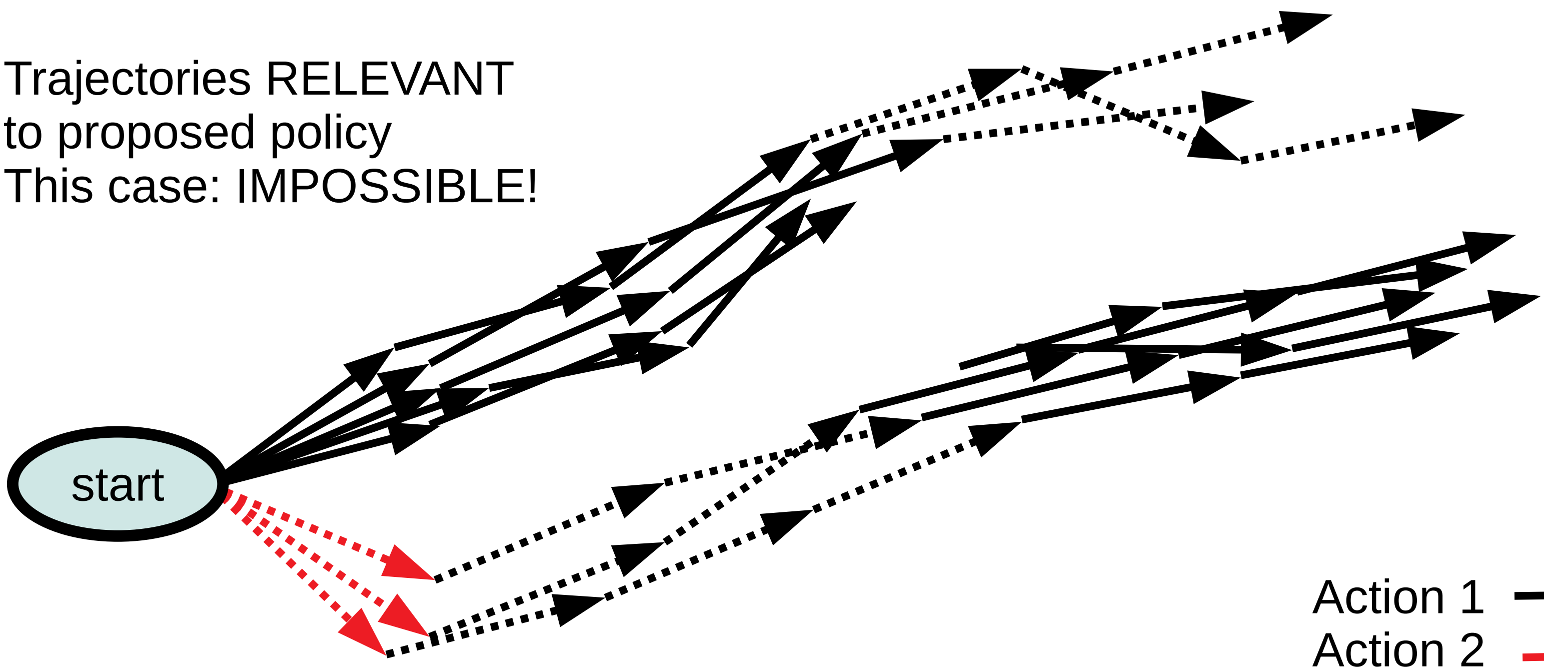
Core idea: Expose sensitive points to humans to validate.



Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

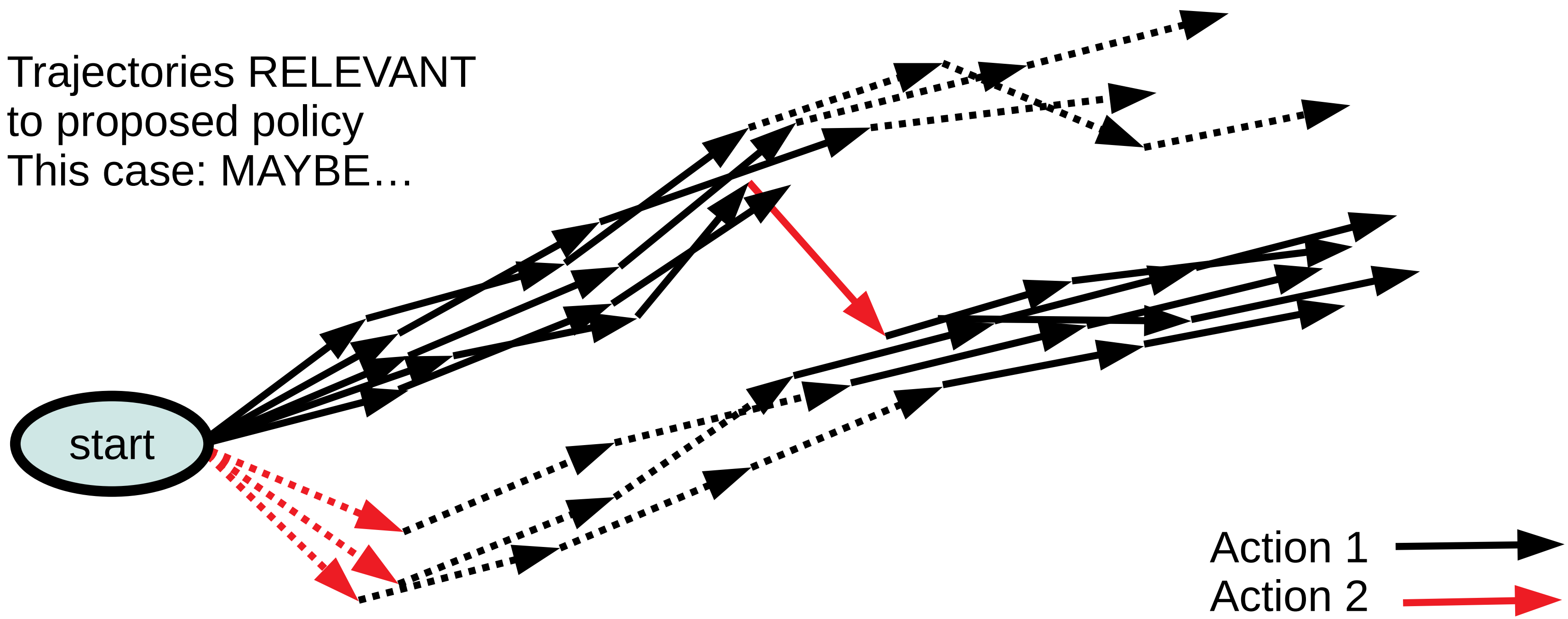
Core idea: Expose sensitive points to humans to validate.



Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

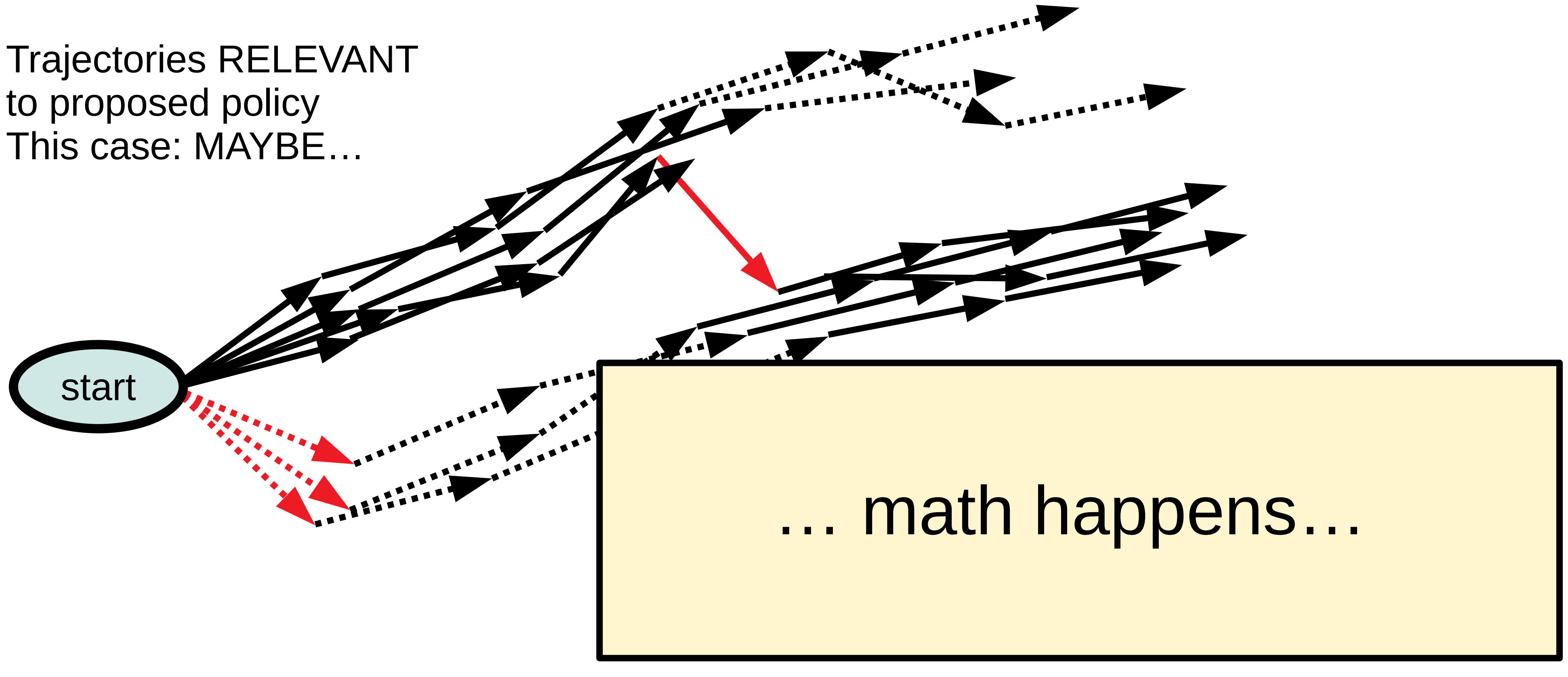
Core idea: Expose sensitive points to humans to validate.



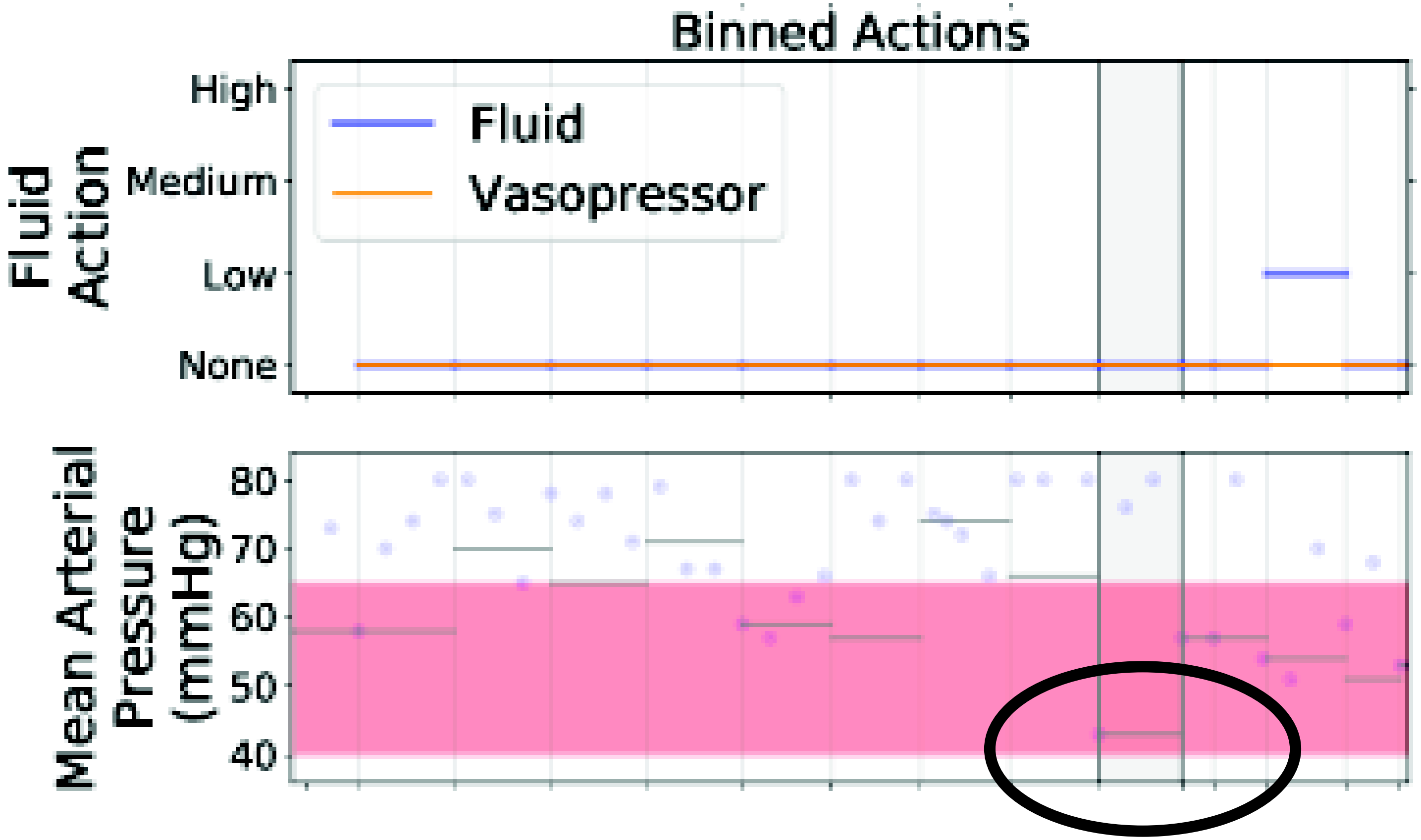
Improving statistical validation with human input

Setting: Estimating the value of a proposed treatment policy.

Core idea: Expose sensitive points to humans to validate.

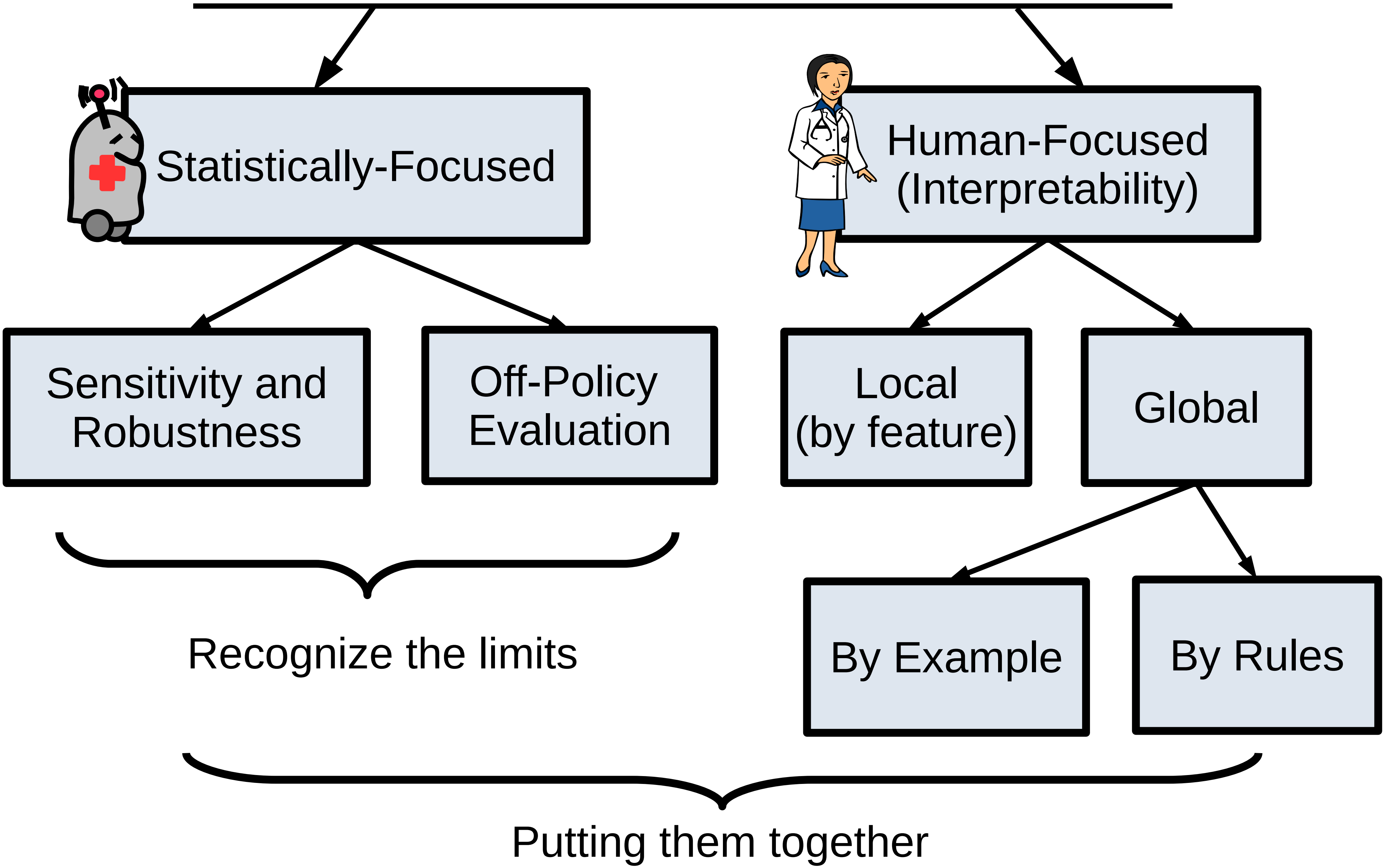


Real Data Example (MIMIC)

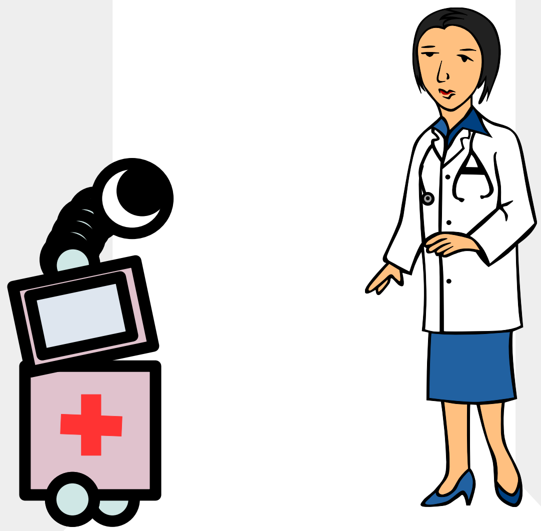


low MAP is a bad reading

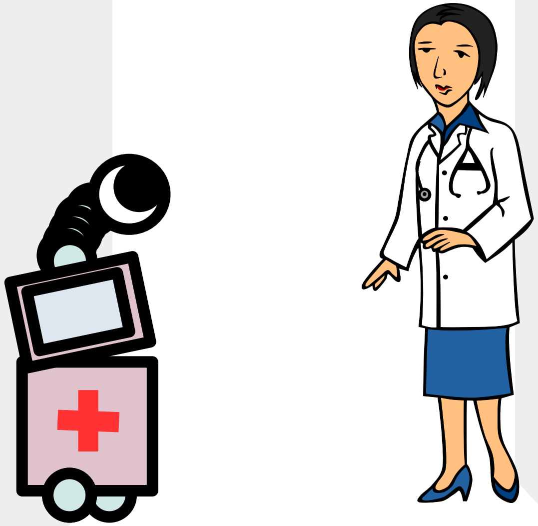
Batch Validation Roadmap



For RL to make an impact in healthcare (and other areas), it’s important to take a holistic approach to validation from the start.

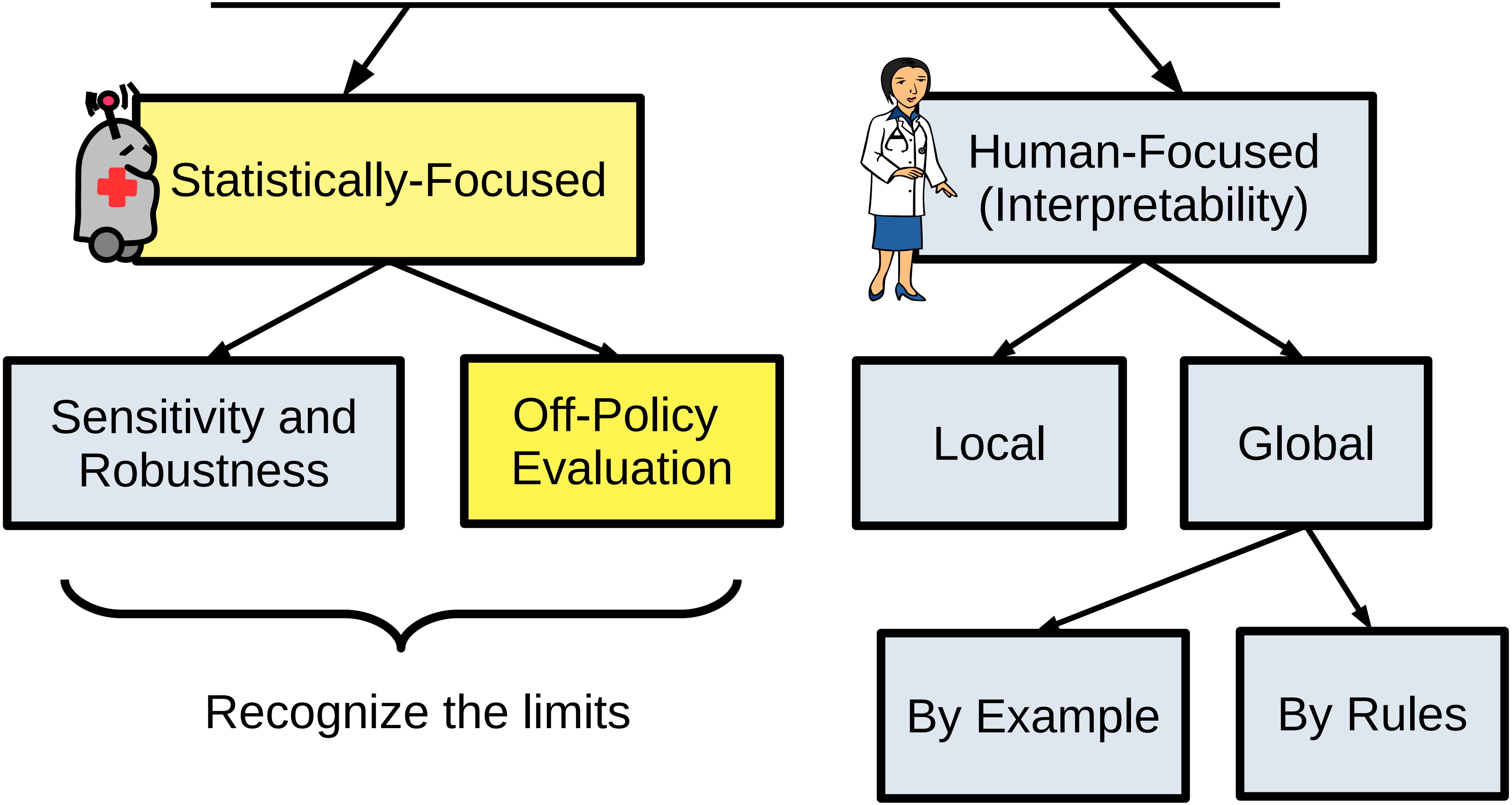


For RL to make an impact in healthcare (and other areas), it's important to take a holistic approach to validation from the start.



Would not be possible without: **DtAK and DtAK alums**: Weiwei Pan, Sonali Parbhoo, Melanie Pradier, Joe Futoma, Michael Hughes, Madhi Pakdaman, Ike Lage, Andrew Ross, Yaniv Yacoby, Jiayu Yao, Beau Coker, Anna Li, Sarah Rathnam, Abhishek Sharma, Eura Shin, Omer Gottesman, Muhammad Arjumand Masood; **Collaborators**: Roy Perlis, Tom McCoy, Taylor Killian, Soumya Ghosh, Xuefeng Peng, David Wihl, Yi Ding, Liwei Lehman, Matthieu Komorowski, Aldo Faisal, David Sontag, Fredrik Johansson, Leo Celi, Aniruddh Raghu, Yao Liu, Emma Brunskill, Sam Gershman, Been Kim, Menaka Narayanan, Emily Chen, Jeffrey He, Ofra Amir, and the CS282 2017; **Admins**: Meg Hastings, Michaela Kapp, Jenny Mileski, Ashley Bens, Annalee Mendez, Jill Sussery, Jasmin Ware, Joanne Bourgeois… and **many, many more** supporters and students at SEAS and beyond!

Batch Validation Roadmap



Off-Policy Evaluation

Core question: Given data collected under some behavior policy π_b, can we estimate the value of some other evaluation policy π_e?

Three main kinds of approaches:

- Importance-sampling: reweight current data (high variance)

$$\rho_n = \prod_t \frac{\pi_e(a_{tn}|s_{tn})}{\pi_b(a_{tn}|s_{tn})}$$

- Model-based: build model with current data, simulate (high bias)
- Value-based: apply value evaluation to current data (high bias)

Off-Policy Evaluation

Core question: Given data collected under some behavior policy π_b, can we estimate the value of some other evaluation policy π_e?

Three main kinds of approaches:

- Importance-sampling: reweight current data (high variance)

$$\rho_n = \prod_t \frac{\pi_e(a_{tn} | s_{tn})}{\pi_b(a_{tn} | s_{tn})}$$

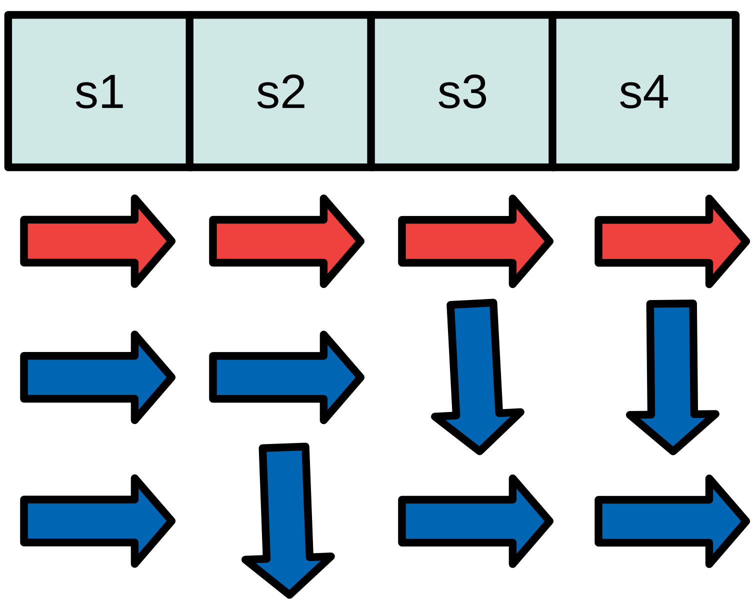
- Model-based: build model with current data, simulate (high bias)
- Value-based: apply value evaluation to current data (high bias)

Stitching to Increase Sample Sizes

Importance sampling-based estimators suffer because importance weights most importance weights get small very fast:

$$\rho_n = \prod_t \frac{\pi_e(a_{tn}|s_{tn})}{\pi_b(a_{tn}|s_{tn})}$$

One way to ameliorate the issue: “stitch” trajectories with zero weight to get more non-zero weight trajectories.



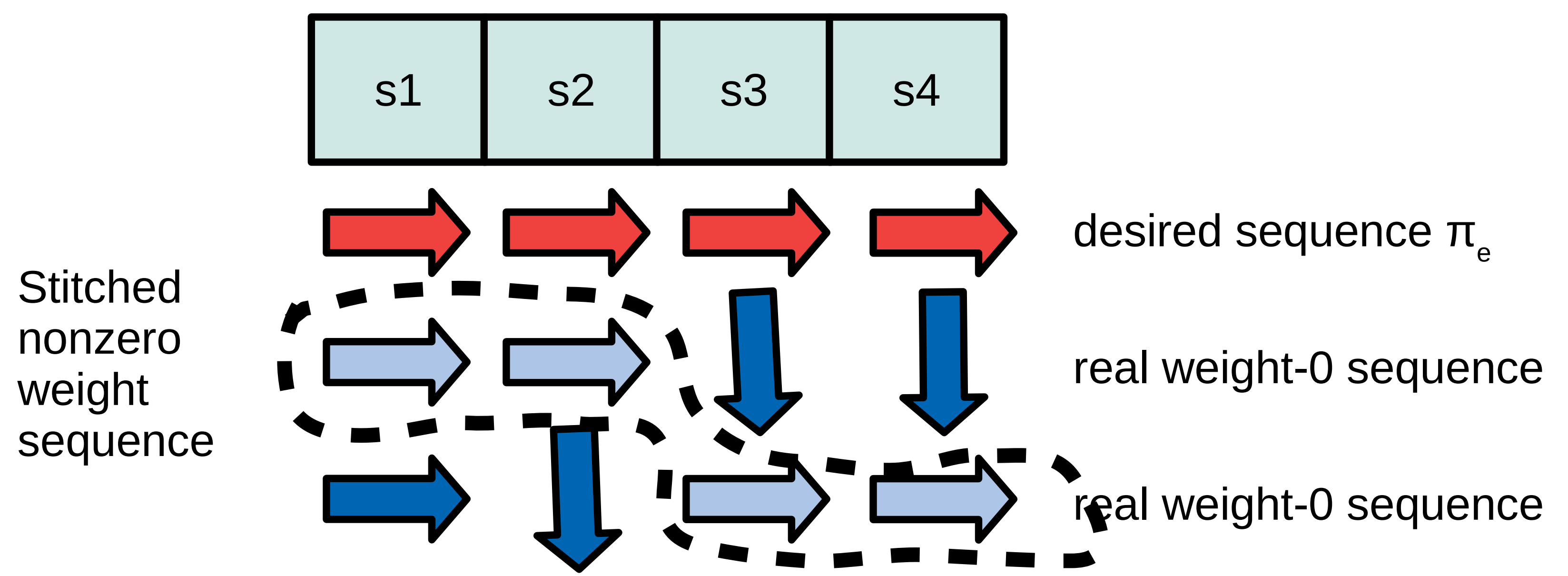
(Sussex et al, ICMLWS 2018)

Stitching to Increase Sample Sizes

Importance sampling-based estimators suffer because importance weights most importance weights get small very fast:

$$\rho_n = \prod_t \frac{\pi_e(a_{tn}|s_{tn})}{\pi_b(a_{tn}|s_{tn})}$$

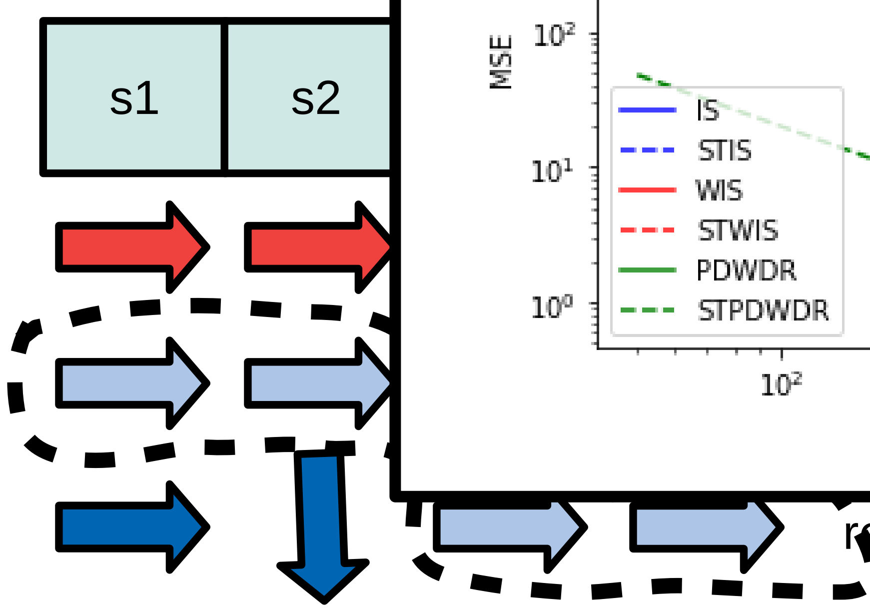
One way to ameliorate the issue: “stitch” trajectories with zero weight to get more non-zero weight trajectories.

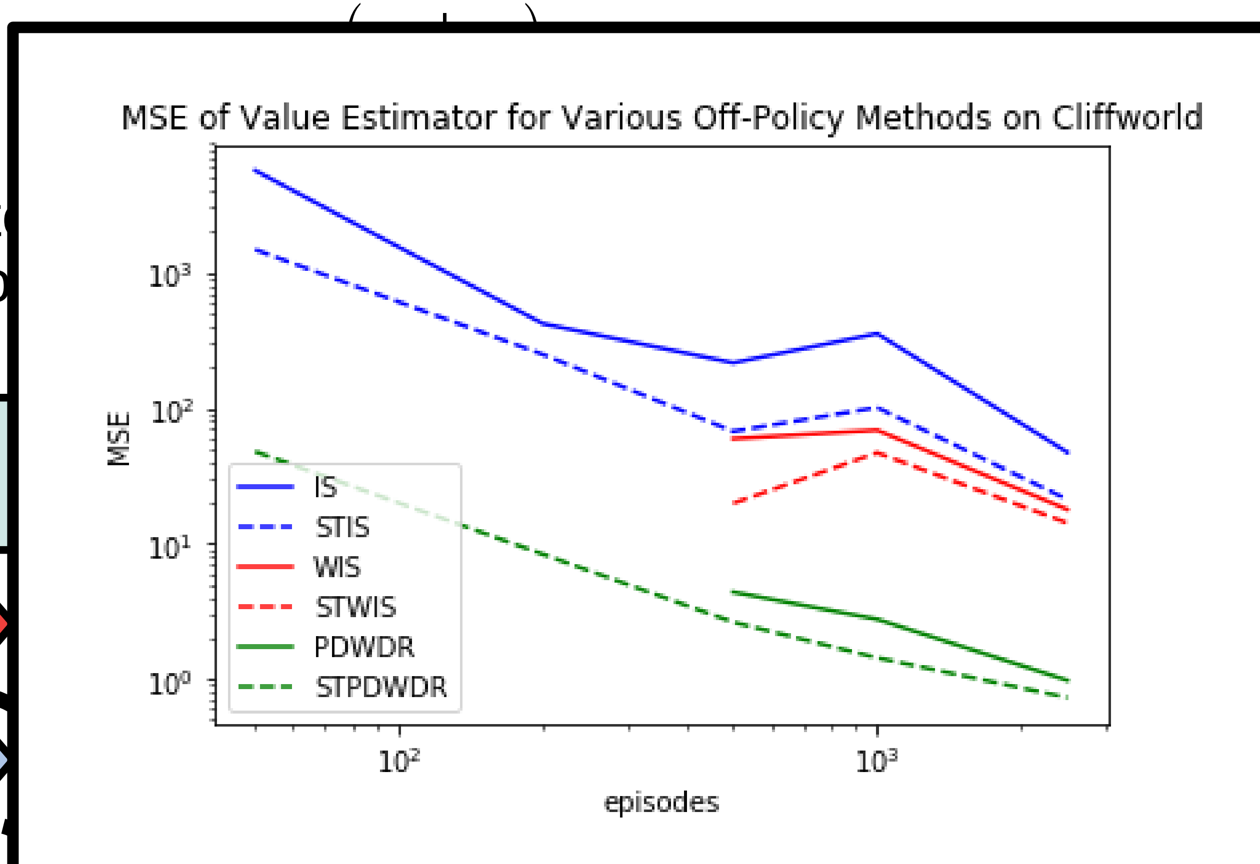


Stitching to Increase Sample Sizes

Importance sampling-based estimators suffer because importance weights most importance weights get small very fast:

One way to ameliorat
weight to get more no





Off-Policy Evaluation

Core question: Given data collected under some behavior policy π_b, can we estimate the value of some other evaluation policy π_e?

Three main kinds of approaches:

- Importance-sampling: reweight current data (high variance)

$$\rho_n = \prod_t \frac{\pi_e(a_{tn} | s_{tn})}{\pi_b(a_{tn} | s_{tn})}$$

- Model-based: build model with current data, simulate (high bias)
- Value-based: apply value evaluation to current data (high bias)

Better Models: Designed for Evaluation

Main objective: find a model that will minimize error in individual treatment effects:

$$\cancel{\left(E_{s_0}\left[V^{\pi}(s_0)\right]-E_{s_0}\left[\hat{V}^{\pi}(s_0)\right]\right)^2}$$

$$E_{s_0}\left[\left(V^{\pi}(s_0)-\hat{V}^{\pi}(s_0)\right)^2\right]$$

where the value function is estimated via trajectories from an approximated model M. Question: Can we do better than just optimizing M for p(M|data)?

Show this can be optimized via a transfer-learning type objective:

$$L(M)=\underbrace{\sum_{nt} l(M,n,t)}_{\text{“on-policy” loss}}+\underbrace{\sum_{nt} \rho_{nt}\, l(M,n,t)}_{\text{“reweighted for } \pi_e\text{” loss}}+\ldots$$

(Liu et al, NIPS 2018)

Better Models: Designed for Evaluation

Main objective: find a model that will minimize error in individual treatment effects:

$$\cancel{\left(E_{s_0}[V^{\pi}(s_0)] - E_{s_0}[\hat{V}^{\pi}(s_0)]\right)^2}$$

$$E_{s_0}\left[\left(V^{\pi}(s_0) - \hat{V}^{\pi}(s_0)\right)^2\right]$$

where the value
approximated m
optimizing M for

Show this can be

$$L(M)$$

Table 1: Root MSE for Cart Pole

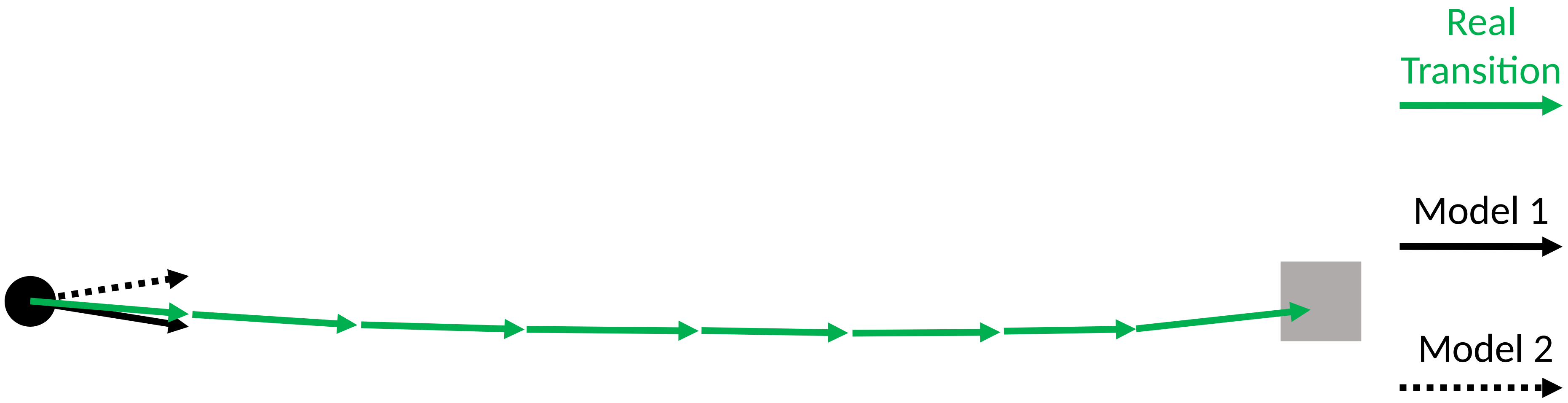
Long Horizon	RepBM	DR	AM	DR(AM)	AM(π)	MRDR Q	MRDR	IS
Mean	**0.4121**	1.359	0.7535	1.786	41.80	151.1	202	194.5
Individual	**1.033**	-	1.313	-	47.63	151.9	-	-
Short Horizon	**RepBM**	**DR**	**AM**	**DR(AM)**	**AM(π)**	**MRDR Q**	**MRDR**	**IS**
Mean	0.07836	**0.02081**	0.1254	0.0235	0.1233	3.013	0.258	2.86
Individual	**0.4811**	-	0.5506	-	0.5974	3.823	-	-

Table 2: Root MSE for Mountain Car

	RepBM	DR	AM	DR(AM)	AM(π)	MRDR Q	MRDR	IS
Mean	**12.31**	135.8	17.15	141.6	72.61	135.4	172.7	149.7
Individual	**31.38**	-	36.36	-	79.46	138.1	-	-

Combining Models

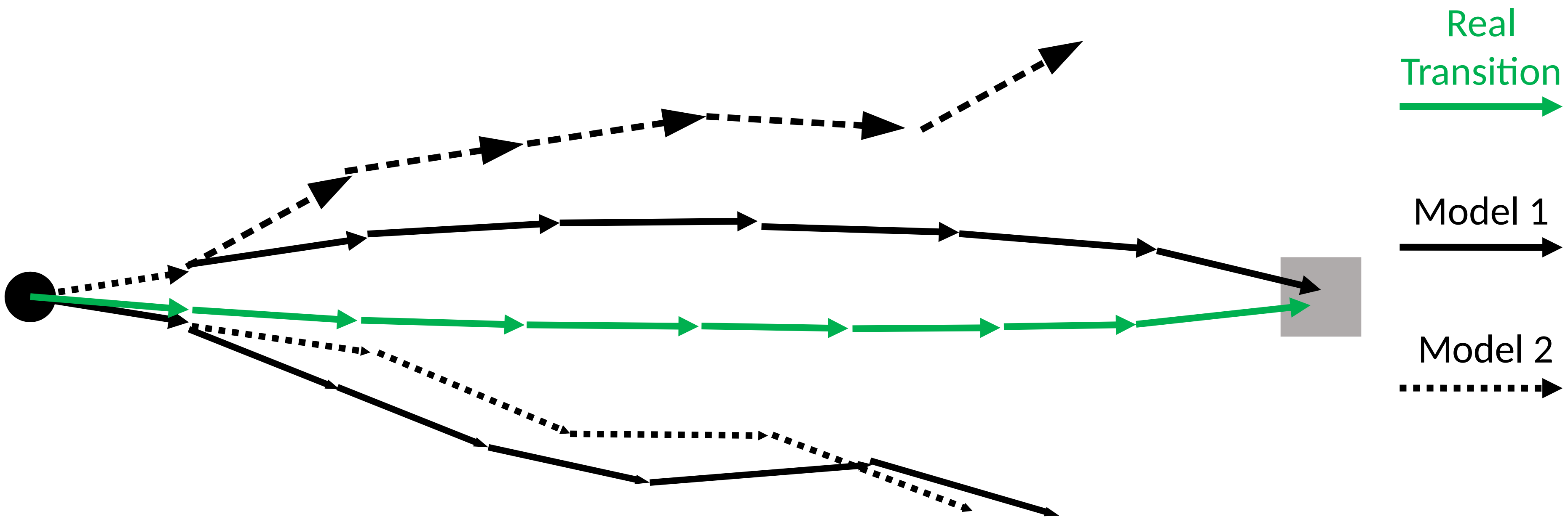
We use RL to bound the long-term accuracy of the value estimate.



(Gottesman et al, ICML 2019)

Combining Models

We use RL to bound the long-term accuracy of the value estimate.



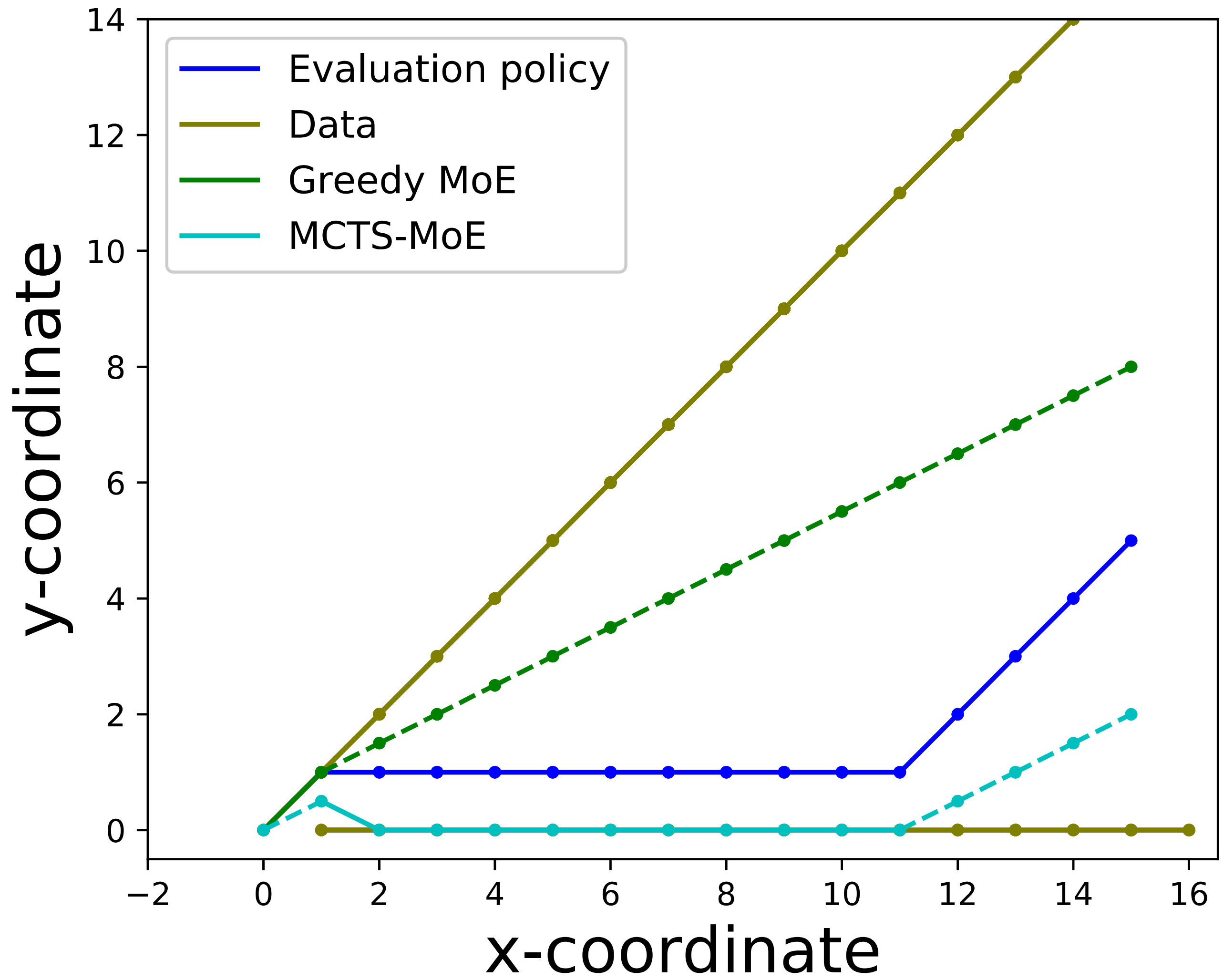
Bound on the Quality

$$\underbrace{\left|g_T - \hat{g}_T\right|}_{\text{Total return error}} \leq \underbrace{L_r \sum_{t=0}^{T} \gamma^t \sum_{t'=0}^{t-1} \left(L_t\right)^{t'} \varepsilon_t\left(t - t' - 1\right)}_{\text{Error due to state estimation}} + \underbrace{\sum_{t=0}^{T} \gamma^t \varepsilon_r(t)}_{\text{Error due to reward estimation}}$$

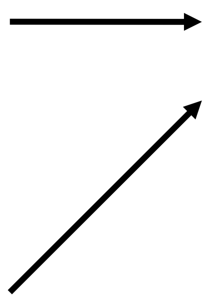
$L_{t/r}$ - Lipschitz constants of transition/reward functions
$\varepsilon_{t/r}(t)$ - Bound on model errors for transition/reward at time t
T - Time horizon
γ - Reward discount factor
$g_T \equiv \sum_{t=0}^{T} \gamma^t r(t)$ - Return over entire trajectory

Closely related to bound in - *Asadi, Misra, Littman. "Lipschitz Continuity in Model-based Reinforcement Learning." (ICML 2018).*

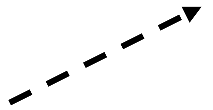
Toy Example



Possible actions

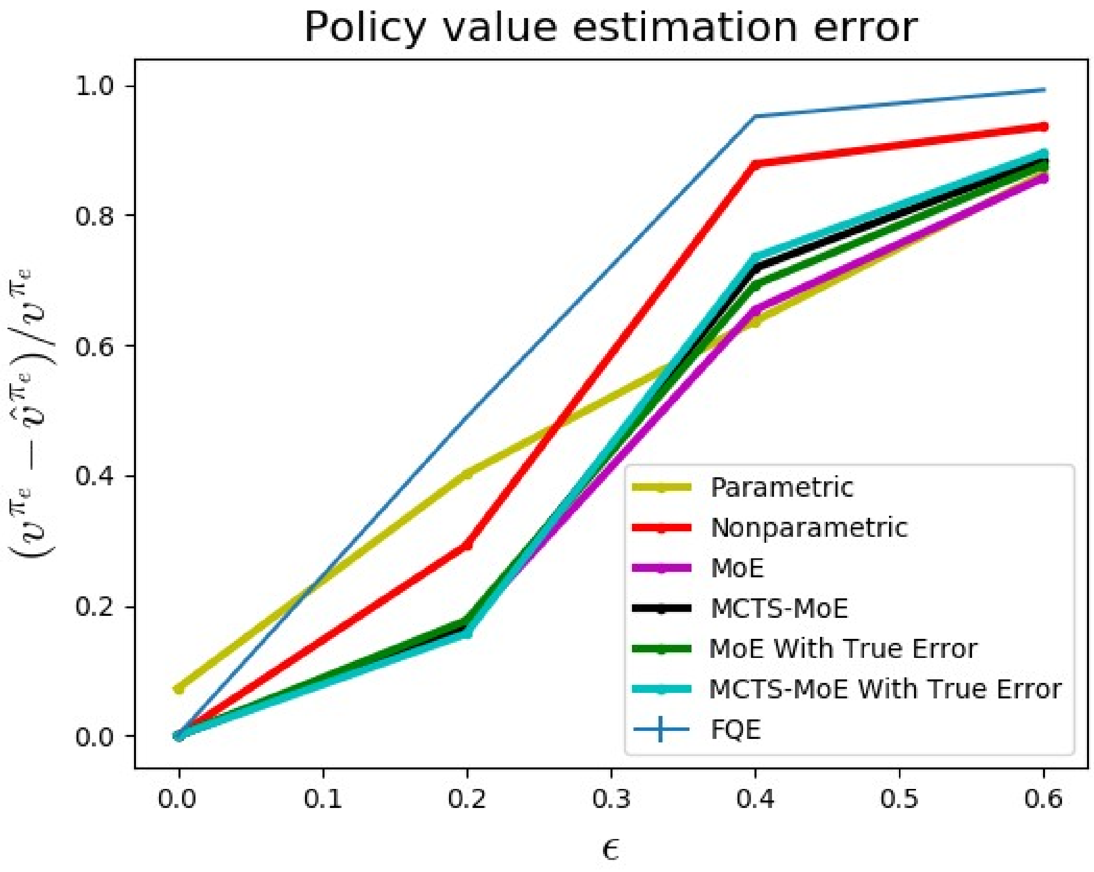


Parametric model



Example with HIV Simulator

We use RL to bound the long-term accuracy of the value estimate.

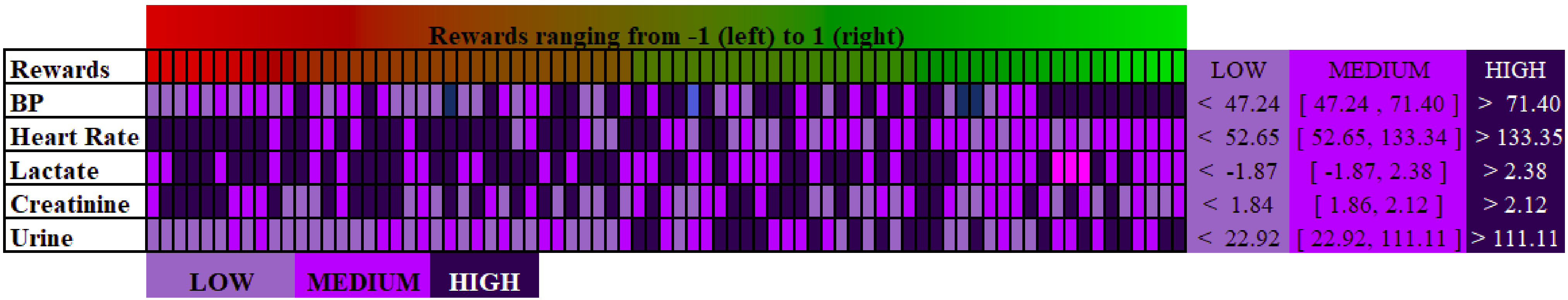


Reward Functions

Helping form reward functions

Reward design is a challenging task for humans. RL can

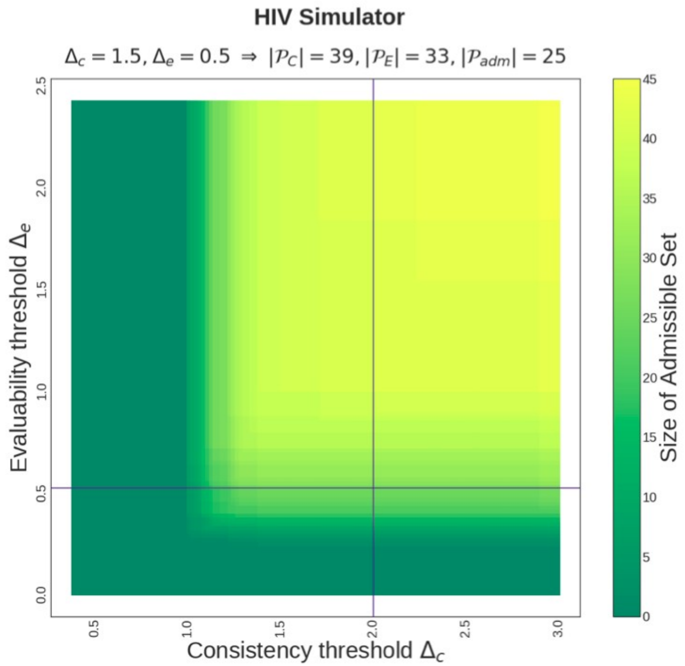
- Extract interpretable rewards that correspond to current behavior for humans to modify (Srinivasan et al. 2020).

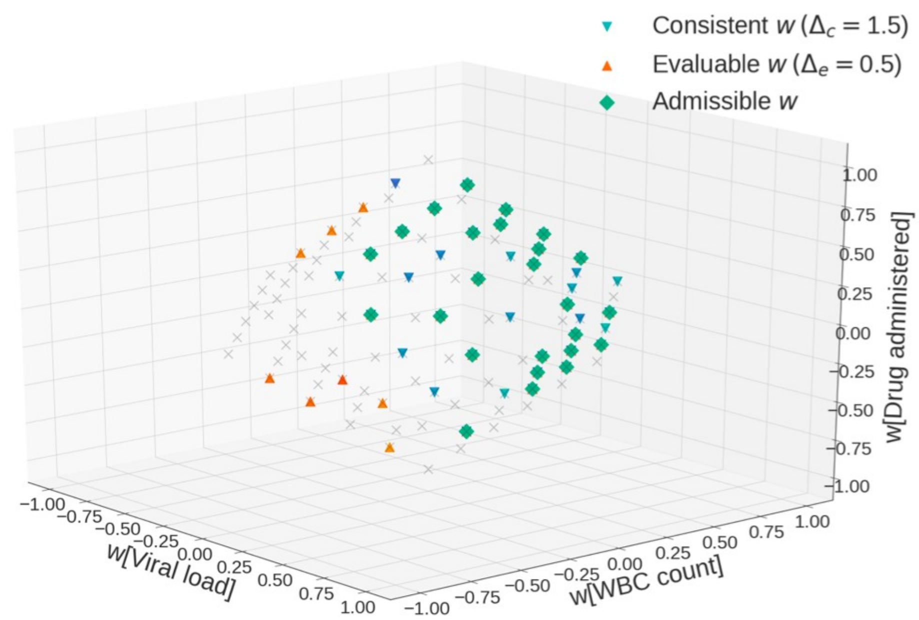


Helping form reward functions

Reward design is a challenging task for humans. RL can

- Extract interpretable rewards that correspond to current behavior for humans to modify (Srinivasan et al. 2020).
- Identify rewards that are consistent with human behavior (Prasad et al. 2020).

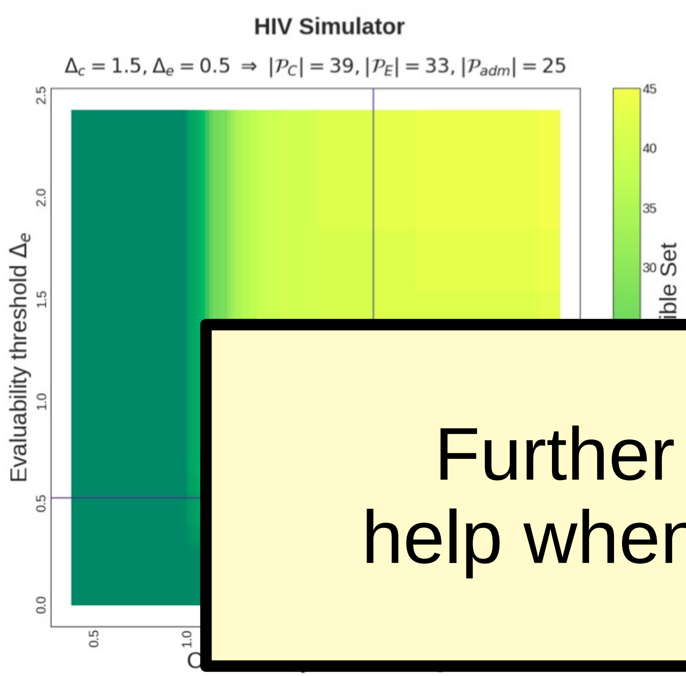


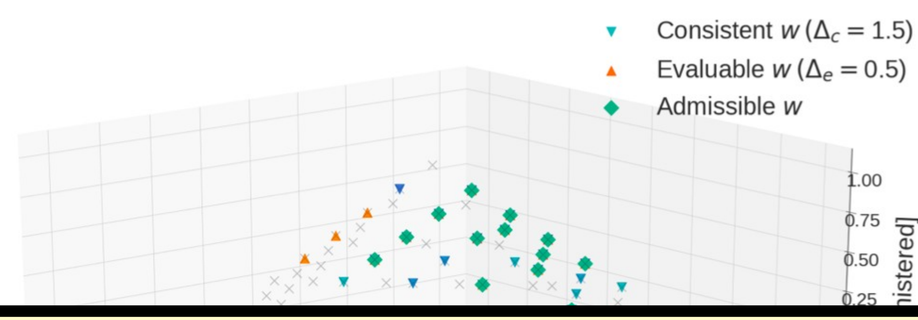


Helping form reward functions

Reward design is a challenging task for humans. RL can

- Extract interpretable rewards that correspond to current behavior for humans to modify (Srinivasan et al. 2020).
- Identify rewards that are consistent with human behavior (Prasad et al. 2020).

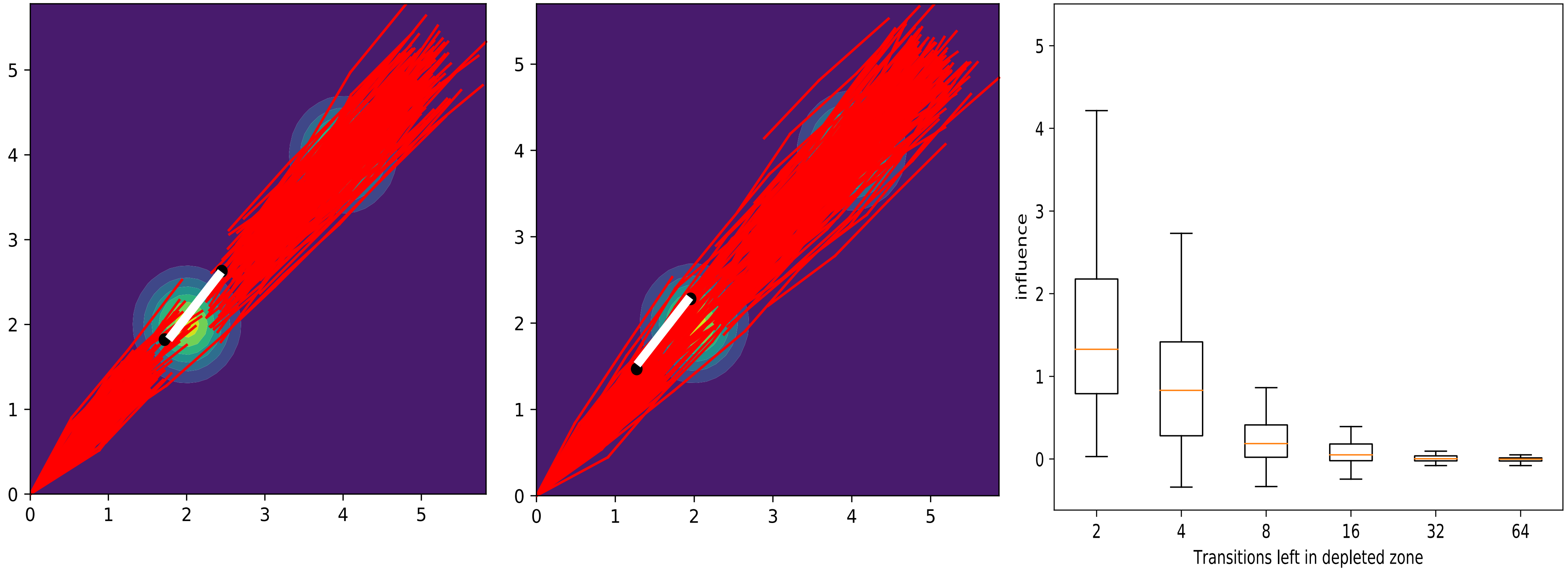




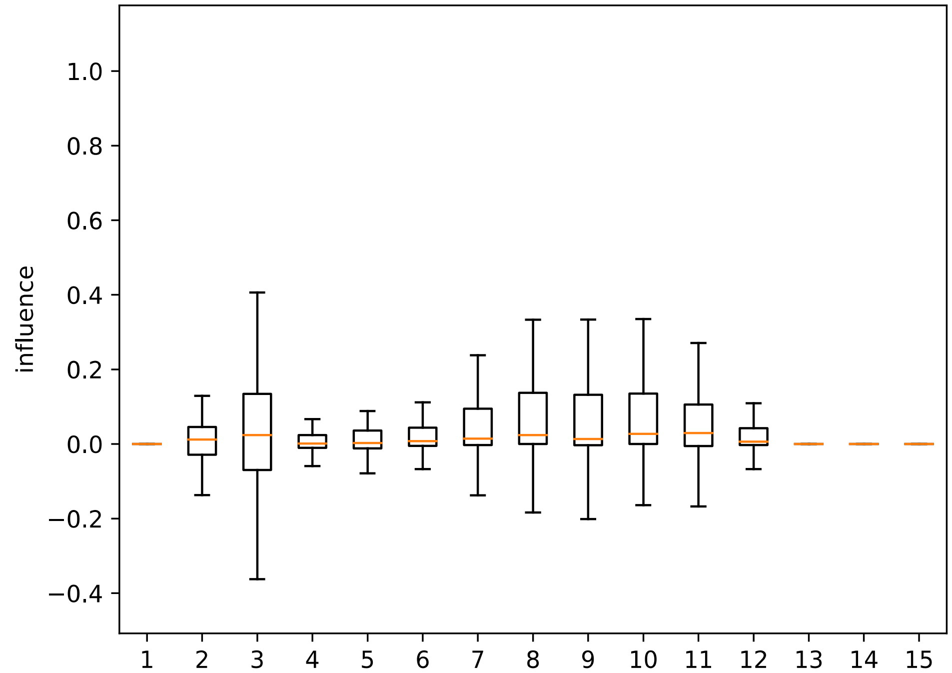
Further directions: How can agents help when rewards are hard to specify?

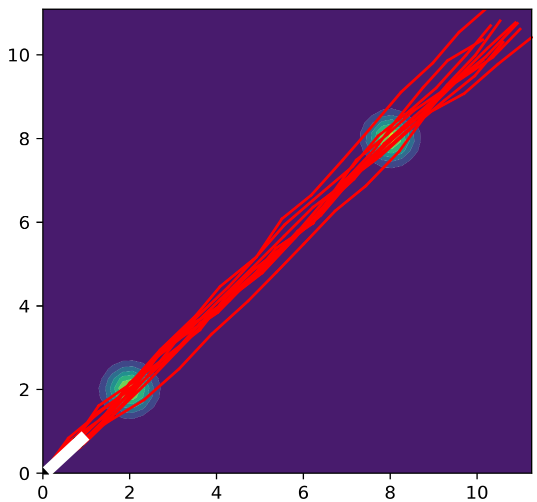
Combining

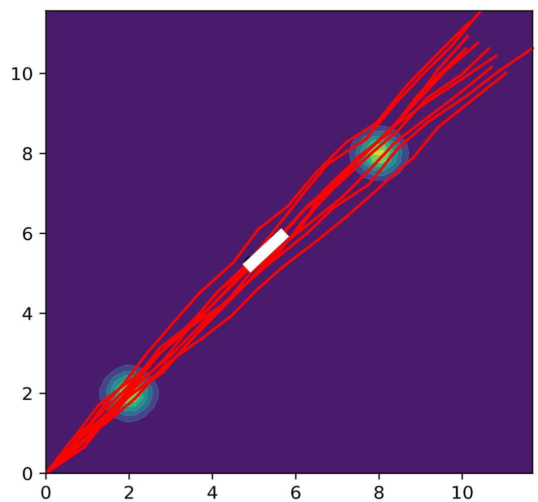
Demonstration on Simple Domains: Influence Depends on Density

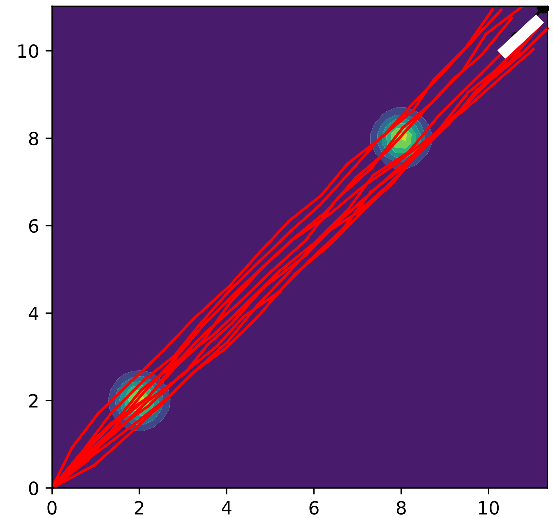


Demonstration on Simple Domains: Influence Depends on Rewards



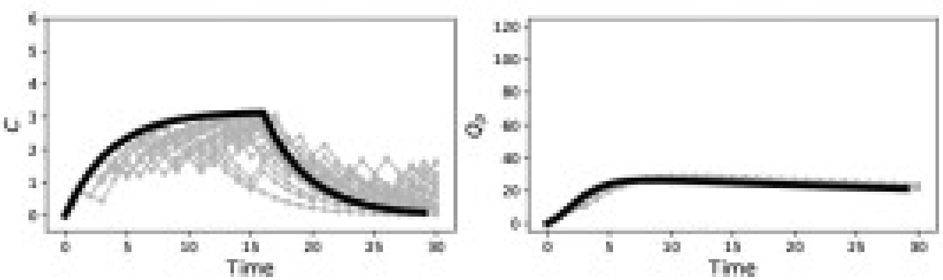




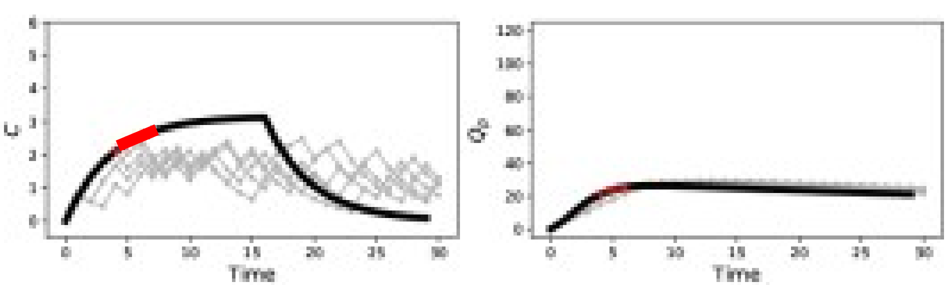


Four Main Cases
(with a 5D cancer simulator)

Stats can determine

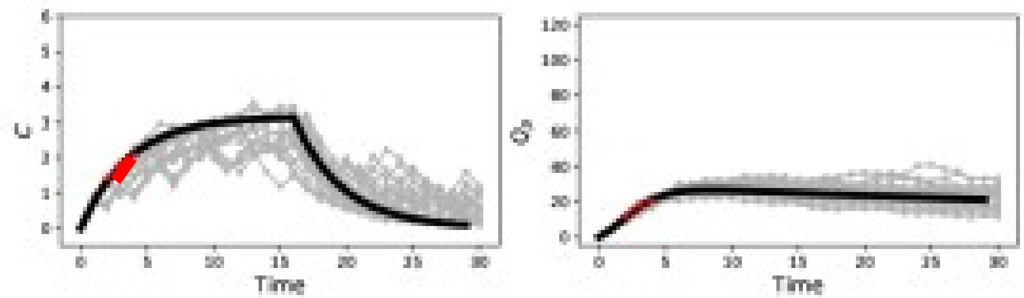


GOOD: No transitions are influential!

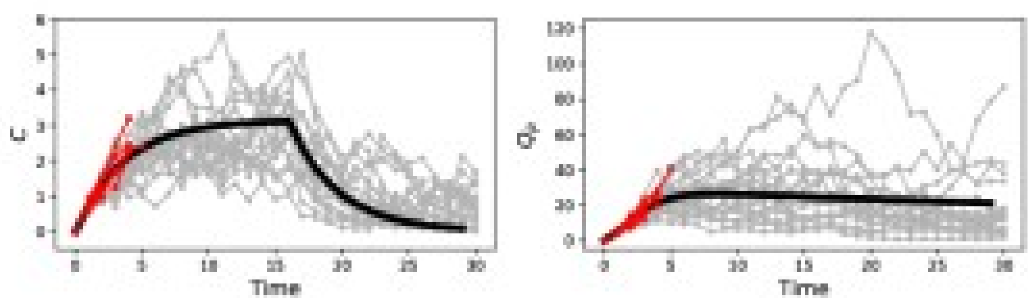


BAD: Influential transition is a “dead end:” no nearest neighbors to continue

Needs a human



GOOD: Influential transition is typical



BAD: Influential transition is not typical

Local Checks

Experts Check Specific Choices

- Example (HIV): check against standard of care

	NNRTIs	NRTIs	PIs	Fusion/Entry Inhibitors
First-line therapy	12 157	3 054	774	128
Second-line therapy	4 068	8 764	6 082	1 042

- Example (HIV): Ask panel of experts

	Clinician 1	Clinician 2	Clinician 3
Agree	18	15	13
Partially Agree	10	11	13
Disagree	2	4	4

Experts Check Specific Choices

- Example (HIV): check against standard of care

	NNRTIs	NRTIs	PIs	Fusion/Entry Inhibitors
First-line therapy	12 157	3 054	774	128
Second-line therapy	4 068	8 764	6 082	1 042

- Example (HIV): Ask panel of experts

Agree
Partially Ag
Disagree

Concern: How do you know if you've checked enough examples?

Maia’s Study

Baseline Approach: Summary Only

Patient Details:

Patricia is a 31 year old woman who is married and works full time. She has a history of seizure disorder and lack of appetite, and presents with 11 months of depressed mood. Current medications include Omeprazole and Celecoxib. Prior treatment with Citalopram did not cause a reduction of depression symptoms.

Baseline Approach: Summary + Rec

Patient Details:

Jennifer is a 40 year old woman who is married and works from home. She is diabetic and has a history of hypertensive heart disease and arrhythmia. She presents with 10 months of depressed mood. Current medications include amoxicillin, and prior treatment with Paroxetine had no effect on depressed mood.

System.07 Recommendation: **DULOXETINE**

Top 5 therapies with highest probability for stability:

Therapy	Predicted Stability*	Predicted Dropout Risk**
Duloxetine	.79	.03
Fluoxetine	.65	.03
Citalopram	.60	.06
Escitalopram	.59	.07
Bupropion	.55	.12

**Stability: continued use of the same medication for at least 3 months*

***Dropout: early treatment discontinuation following prescription*

Approach: Placebo Explanation

Patient Details:

David is a 43 year old man who is widowed and works full time. He has a history of diabetes, arrhythmia and hypertensive heart disease. He presents with 14 months of depressed mood. Current medications include amoxicillin, and prior treatment with Paroxetine was ineffective.

System.10 Recommendation: **DULOXETINE**

Top 5 therapies with highest probability for stability:

Therapy	Predicted Stability*	Predicted Dropout Risk**
Duloxetine	.78	.06
Fluoxetine	.65	.08
Citalopram	.54	.10
Escitalopram	.52	.14
Bupropion	.52	.14

**Stability: continued use of the same medication for at least 3 months*

***Dropout: early treatment discontinuation following prescription*

Why are these therapies being recommended?

System.10's predictions are **based on the patient's ICD-9 codes.**

Discovery: Last decade+ of objectives are label replication!

- Posterior regularization: Choose $q(\theta_n)$ to optimize a lower bound on $p(x_n|\varphi)$ with some constraint, for example, $E_q[\text{loss}(y_n,\hat{y}_n)]$ bounded.

$$-\sum_n E_{q(\theta_n)}[\log p(x_n,\theta_n|\phi)+\lambda\log p(y_n|x_n,\theta_n,\eta)-\log q(\theta_n)]$$

- MED-LDA/Regularized Bayesian Inference introduces a margin constraint.

$$-\sum_n L(x_n,q,\phi)+C\sum_n \log p(E_{q(\theta_n)}[\hat{y}_n])$$

More ways to get small models

- Nearly Simple Models: Make model close to a decision tree.
- D ... ere
  cl

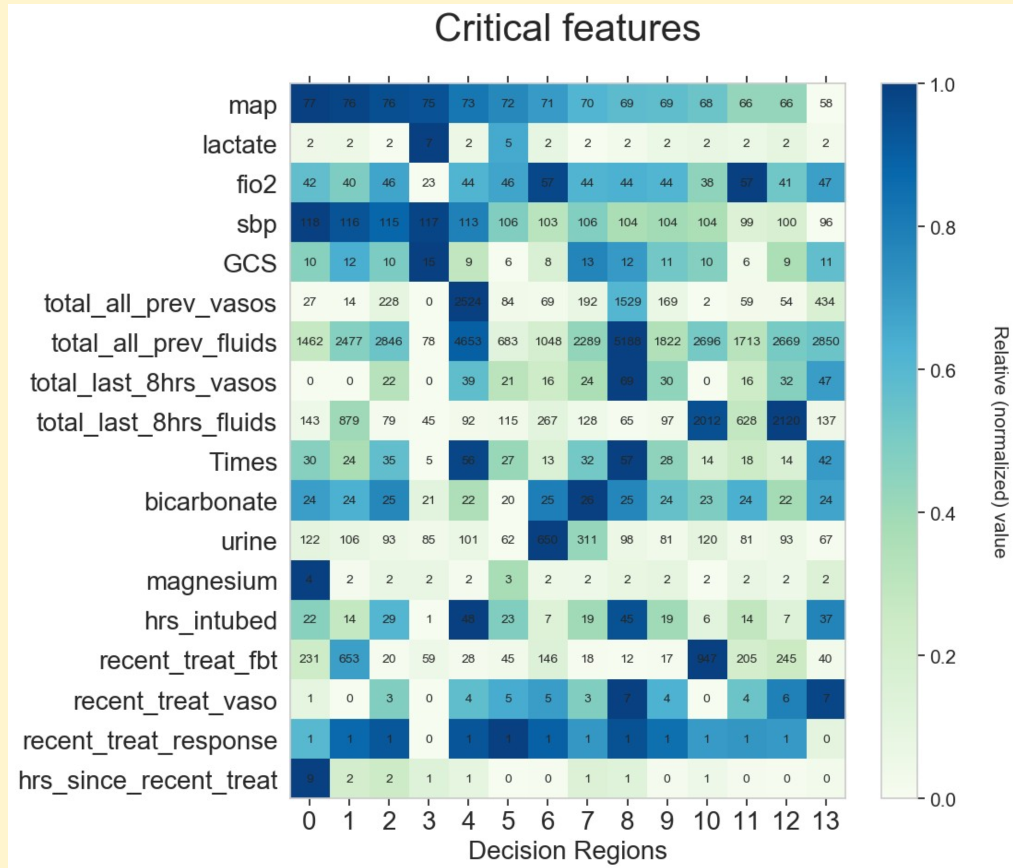
Pati

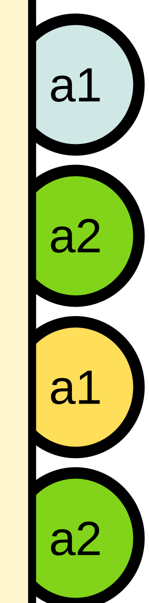
Pati

Pati

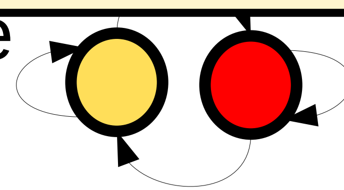
Pati

Back to Hypotension: State Summaries





Just two states to optimize; we can build a tiny 2-state MDP!



Zhang et al. ML4H WS 2020