Multiagent Systems, Reinforcement Learning, and Robotics

Peter Stone

Learning Agents Research Group (LARG)
Department of Computer Science
The University of Texas at Austin

UT Austin: Exciting Times!

- NSF Institute for Foundations of Machine Learning (IFML)
  - **Machine Learning Laboratory** (MLL)

UT Austin: Exciting Times!

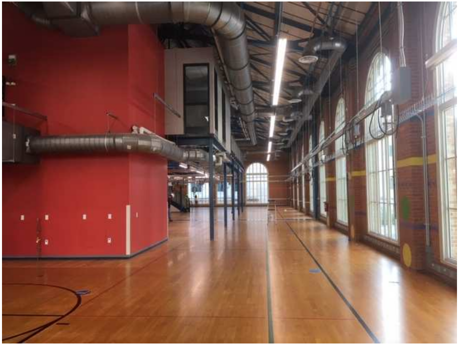
- NSF Institute for Foundations of Machine Learning (IFML)
  - **Machine Learning Laboratory** (MLL)
- Bridging Barriers: **Good Systems**

UT Austin: Exciting Times!

- NSF Institute for Foundations of Machine Learning (IFML)
  - **Machine Learning Laboratory** (MLL)
- Bridging Barriers: **Good Systems**
- Director of **Texas Robotics**
  - Ribbon cutting on new building last October

UT Austin: Exciting Times!

- NSF Institute for Foundations of Machine Learning (IFML)
  - **Machine Learning Laboratory** (MLL)
- Bridging Barriers: **Good Systems**
- Director of **Texas Robotics**
  - Ribbon cutting on new building last October




TEXAS
The University of Texas at Austin
WHAT STARTS HERE CHANGES THE WORLD
Texas Robotics Faculty
Peter Stone
Computer Science
Joydeep Biswas
Computer Science
Scott Niekum
Computer Science
Yuke Zhu
Computer Science
Farshid Alambeigi
Mechanical Engineering
Ashish Deshpande
Mechanical Engineering
Ann Majewicz Fey
Mechanical Engineering
Mitch Pryor
Mechanical Engineering
James Sulzer
Mechanical Engineering
Sandeep Chinchali
Electrical & Computer Engineering
Andrea Thomaz
Electrical & Computer Engineering
José del R. Millán
Electrical & Computer Engineering
David Fridovich-Keil
Aerospace Engineering & Engineering Mechanics
Luis Sentis
Aerospace Engineering & Engineering Mechanics
Ufuk Topcu
Aerospace Engineering & Engineering Mechanics

The Big Scientific Questions of our Time

The Big Scientific Questions of our Time

- How did the **universe** originate?

The Big Scientific Questions of our Time

- How did the **universe** originate?
- How did **life** on Earth originate?

The Big Scientific Questions of our Time

- How did the **universe** originate?
- How did **life** on Earth originate?
- What is the nature of **intelligence**?

The Nature of Intelligence

The Nature of Intelligence

How Can we Study it?

The Nature of Intelligence

How Can we Study it?

- Think about it

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior — Psychology

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior — Psychology
- Study human (or other animal) brains

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior — Psychology
- Study human (or other animal) brains — Neuroscience

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior — Psychology
- Study human (or other animal) brains — Neuroscience
- Build and analyze intelligent artifacts

The Nature of Intelligence

How Can we Study it?

- Think about it — Philosophy
- Study human (or other animal) behavior — Psychology
- Study human (or other animal) brains — Neuroscience
- Build and analyze intelligent artifacts — Computer Science

A Goal of AI and Robotics

A Goal of AI and Robotics

Robust, **fully autonomous**
agents in the real world

A Goal of AI and Robotics

$\boxed{\text{Robust, \textbf{fully autonomous} agents in the real world}}$

How?

How to Build Intelligent Agents?

How to Build Intelligent Agents?

Russell, '95

> "Theoreticians can produce the AI equivalent of bricks, beams, and mortar with which AI architects can build the equivalent of cathedrals."

How to Build Intelligent Agents?

Russell, '95

"Theoreticians can produce the AI equivalent of bricks, beams, and mortar with which AI architects can build the equivalent of cathedrals."

Koller, '01

"In AI . . . we have the tendency to divide a problem into well-defined pieces, and make progress on each one.

How to Build Intelligent Agents?

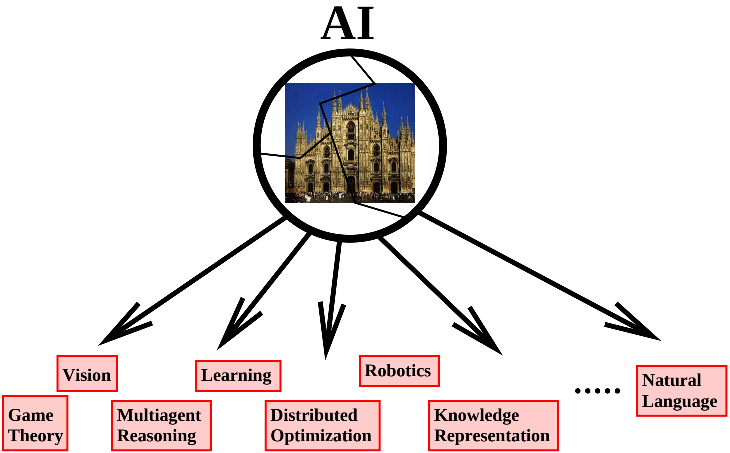
Russell, '95

"Theoreticians can produce the AI equivalent of bricks, beams, and mortar with which AI architects can build the equivalent of cathedrals."

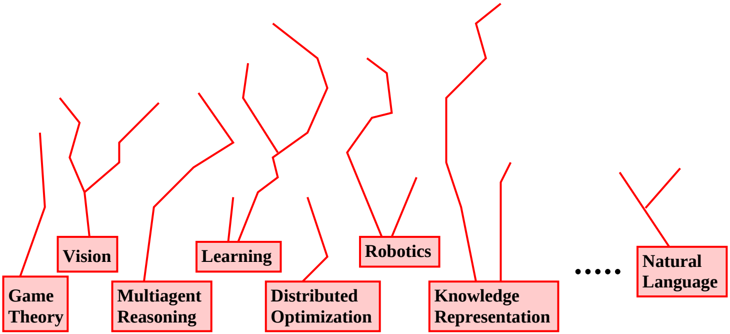
Koller, '01

"In AI . . . we have the tendency to divide a problem into well-defined pieces, and make progress on each one. . . . Part of our solution to the AI problem must involve building bridges between the pieces."

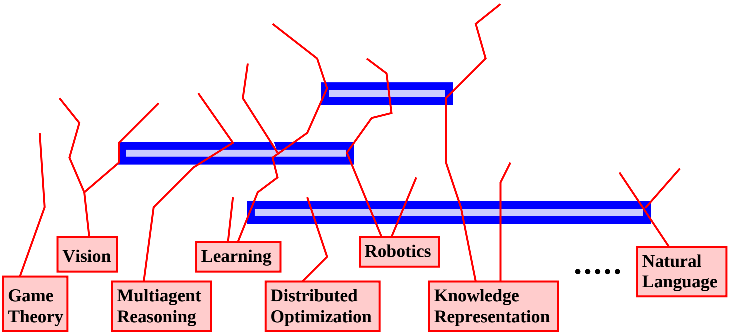
Dividing the Problem



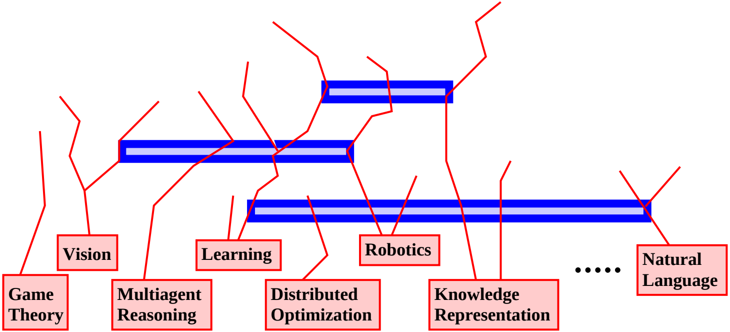
The Bricks


Vision
Learning
Robotics
Game Theory
Multiagent Reasoning
Distributed Optimization
Knowledge Representation
.....
Natural Language

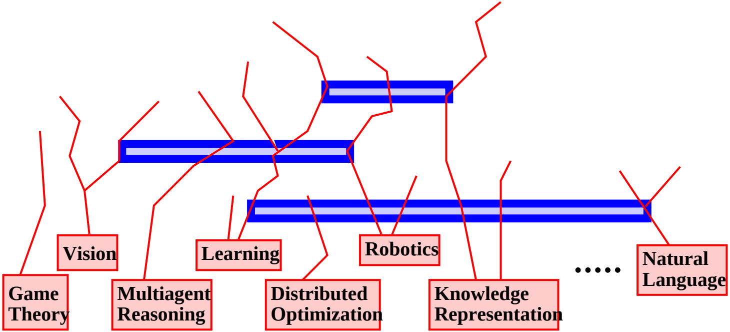
The Beams and Mortar



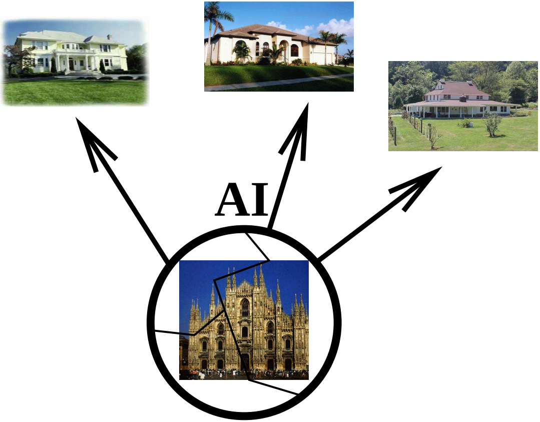
Towards a Cathedral?



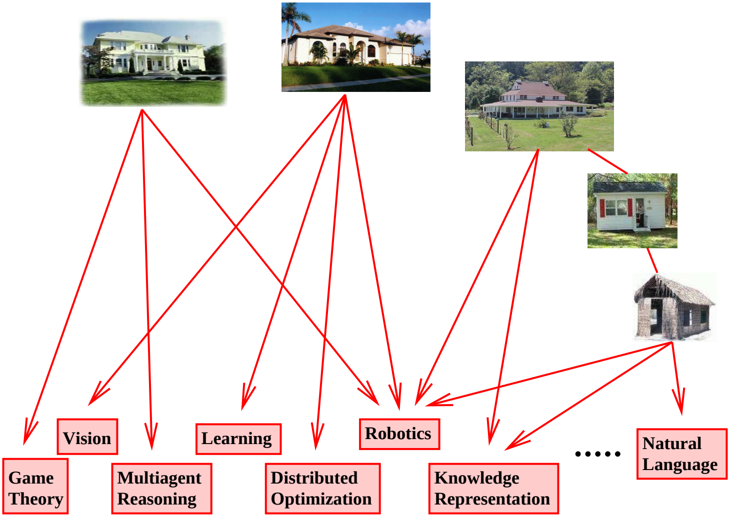
Or Something Else?



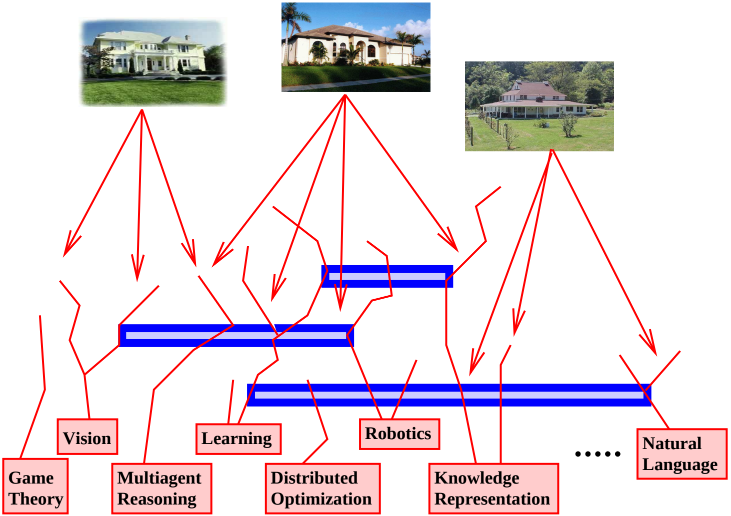
A Different Problem Division



Top-Down Approach



Meeting in the Middle



A Goal of AI and Robotics

Robust, **fully autonomous** agents in the real world

How?

A Goal of AI and Robotics

Robust, **fully autonomous** agents in the real world

How?

- Build complete agents to perform increasingly complex tasks

A Goal of AI and Robotics

Robust, **fully autonomous** agents in the real world

How?

- Build complete agents to perform increasingly complex tasks
  Complete agents: sense, decide, and act — closed loop

A Goal of AI and Robotics

Robust, **fully autonomous**
agents in the real world

How?

- Build complete agents to perform increasingly complex tasks
  Complete agents: sense, decide, and act — closed loop
- Drives research on component algorithms, theory

A Goal of AI and Robotics

Robust, **fully autonomous**
agents in the real world

How?

- Build complete agents to perform increasingly complex tasks
  Complete agents: sense, decide, and act — closed loop

- Drives research on component algorithms, theory
  - Improve from experience (Machine learning)

A Goal of AI and Robotics

Robust, **fully autonomous** agents in the real world

How?

- Build complete agents to perform increasingly complex tasks
  Complete agents: sense, decide, and act — closed loop
- Drives research on component algorithms, theory
  - Improve from experience (Machine learning)
  - Interact with other agents (Multiagent systems)

A Goal of AI and Robotics

Robust, **fully autonomous**
agents in the real world

How?

- Build complete agents to perform increasingly complex tasks
  Complete agents: sense, decide, and act — closed loop
- Drives research on component algorithms, theory
  - Improve from experience (Machine learning)
  - Interact with other agents (Multiagent systems)

"Good problems produce good science"

My Research Problem

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

My Research Problem

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- Robotics

My Research Problem

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

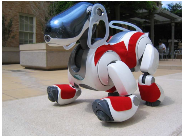
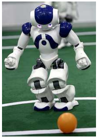
- Autonomous agents
- Multiagent systems
- Robotics
- Machine learning
  - Reinforcement learning

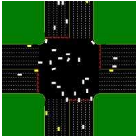
My Research Problem

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- Robotics
- Machine learning
  - Reinforcement learning



RoboCup Soccer

RoboCup Soccer

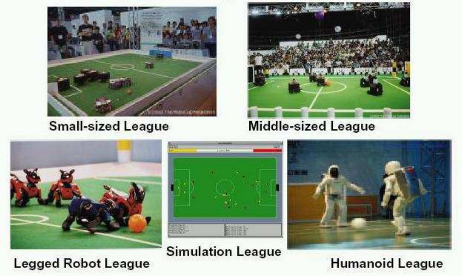
- Grand challenge: beat World Cup champions by 2050

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages

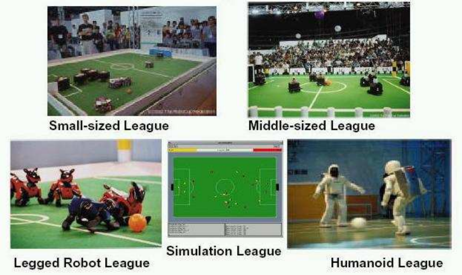
RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many



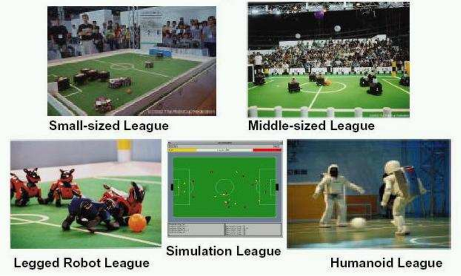
RoboCup Soccer

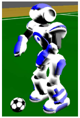
- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  



RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  

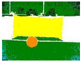




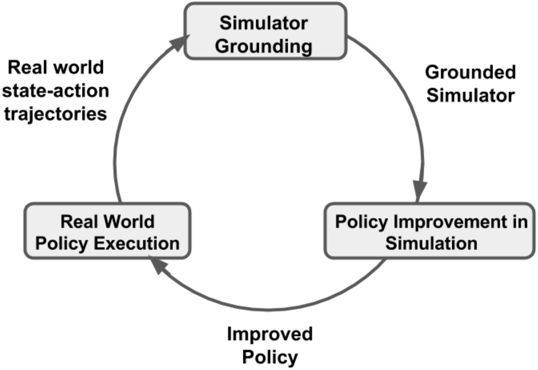
RoboCup@Home

RoboCup@Home

Robot Vision

- Great progress in computer vision
  - Shape modeling, object recognition, face detection. . .
- Robot vision offers new challenges 
  - Mobile camera, limited computation, color features
- **Autonomous color learning** [Sridharan & Stone, '05]
  - Learns color map based on known object locations
  - Recognizes and reacts to illumination changes
  - Object detection in real-time, on-board a robot 

Robot Walking: Grounded Simulation Learning



Method	Velocity (cm/s)	% Improve
Initial policy	19.3	0.0
1st iteration	26.3	34.6
2nd iteration	**28.0**	43.3

RL Research

- Human interaction

RL Research

- Human interaction
  - Advice, Demonstration 

RL Research

- Human interaction
  - Advice, Demonstration 
  - Positive/Negative Feedback

[Knox & Stone, '09]

RL Research

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, '09]

- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]

RL Research

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, '09]

- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]

RL Research

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, ’09]

- Transfer learning for RL [Taylor & Stone, ’07]
- Curriculum Learning [Narvekar et al., ’16]
- RL for musical playlist recommendation [Liebman et al., ’15]
- TEXPLORE for Robot RL [Hester & Stone, ’13]

RL Research

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, '09]

- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]
- TEXPLORE for Robot RL [Hester & Stone, '13]
  - Sample efficient; real-time
  - Continuous state; delayed effects

RL Research

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, '09]

- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]
- TEXPLORE for Robot RL [Hester & Stone, '13]
  - Sample efficient; real-time
  - Continuous state; delayed effects
- Deep RL in continuous action spaces [Hausknecht & Stone, '16]

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]

Multiagent Systems Research

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

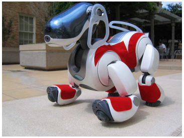
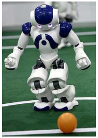
- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]
  - **RoboCup experiments** [Genter et al., '15]

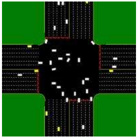
My Research Problem

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- Robotics
- Machine learning
  - Reinforcement learning



Where are we now?

Me: Lots of open research challenges

Where are we now?

Me: Lots of open research challenges

- Professor of Computer Science at UT Austin
- Machine Learning Lab, Good Systems,
- Director of Texas Robotics

Where are we now?

Me: Lots of open research challenges

- Professor of Computer Science at UT Austin
- Machine Learning Lab, Good Systems,
- Director of Texas Robotics

You: Lots of choices and opportunities

Where are we now?

Me: Lots of open research challenges

- Professor of Computer Science at UT Austin
- Machine Learning Lab, Good Systems,
- Director of Texas Robotics

You: Lots of choices and opportunities

- Where will you go to college?
- What will you study?
- What will your lifelong challenge be?

Where are we now?

Me: Lots of open research challenges

- Professor of Computer Science at UT Austin
- Machine Learning Lab, Good Systems,
- Director of Texas Robotics

You: Lots of choices and opportunities

- Where will you go to college?
- What will you study?
- What will your lifelong challenge be?
- Huge opportunity: UTCS, FRI, Turing Scholars

Where are we now?

Me: Lots of open research challenges

- Professor of Computer Science at UT Austin
- Machine Learning Lab, Good Systems,
- Director of Texas Robotics

You: Lots of choices and opportunities

- Where will you go to college?
- What will you study?
- What will your lifelong challenge be?
- Huge opportunity: UTCS, FRI, Turing Scholars

AI: Thriving, but with concerns

A Goal of AI

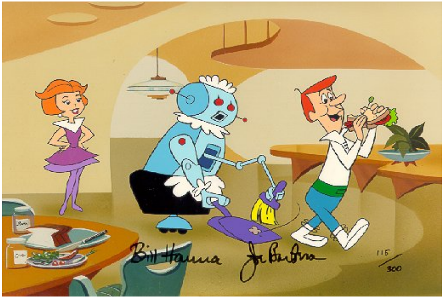
Robust, fully autonomous agents in the real world

What happens **when** we achieve this goal?

A Goal of AI

Robust, fully autonomous agents in the real world

What happens **when** we achieve this goal?

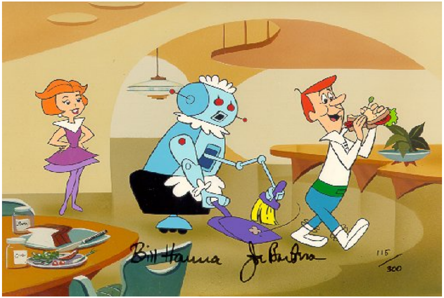


?

A Goal of AI

Robust, fully autonomous agents in the real world

What happens **when** we achieve this goal?



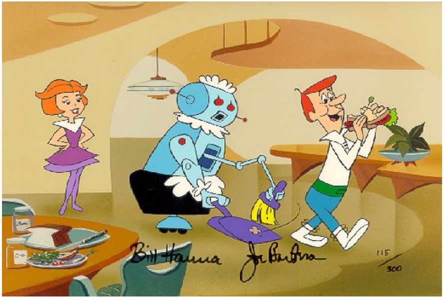
?

?

A Goal of AI

Robust, fully autonomous agents in the real world

What happens **when** we achieve this goal?



?

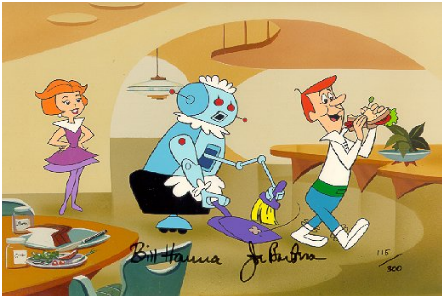
?

- Question: Would you rather have been born
  - 50 years earlier? Or 50 years later?

A Goal of AI

Robust, fully autonomous agents in the real world

What happens **when** we achieve this goal?



?

?

- Question: Would you rather have been born
  - 50 years earlier? Or 50 years later?
- Not clear — world changing in many ways for the worse

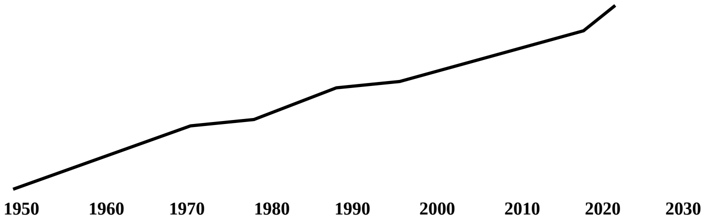
AI can be a part of the solution

Multiagent Systems, Reinforcement Learning, and Robotics

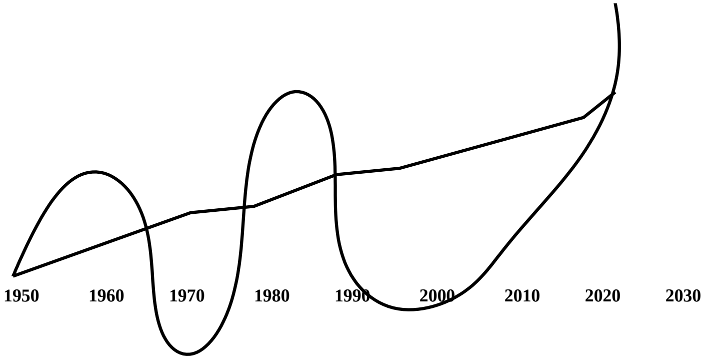
Peter Stone

Learning Agents Research Group (LARG)
Department of Computer Science
The University of Texas at Austin

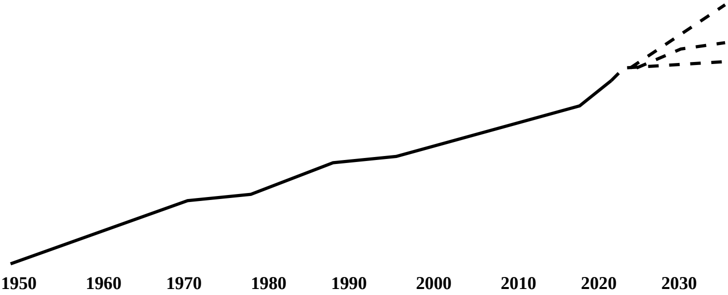
Reality



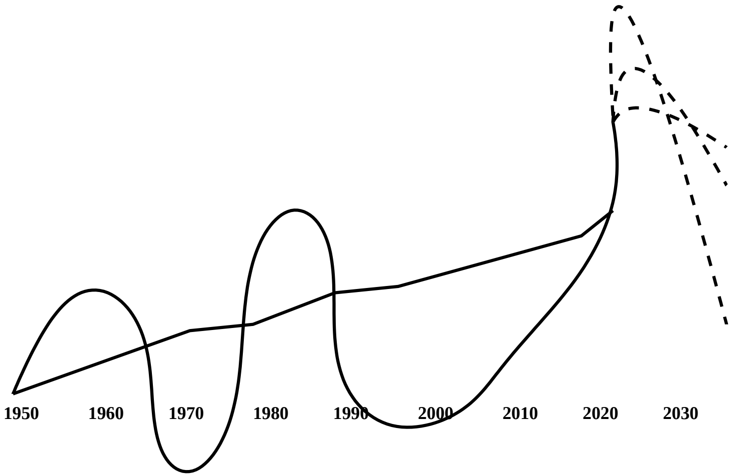
Perceptions



Uncertainty



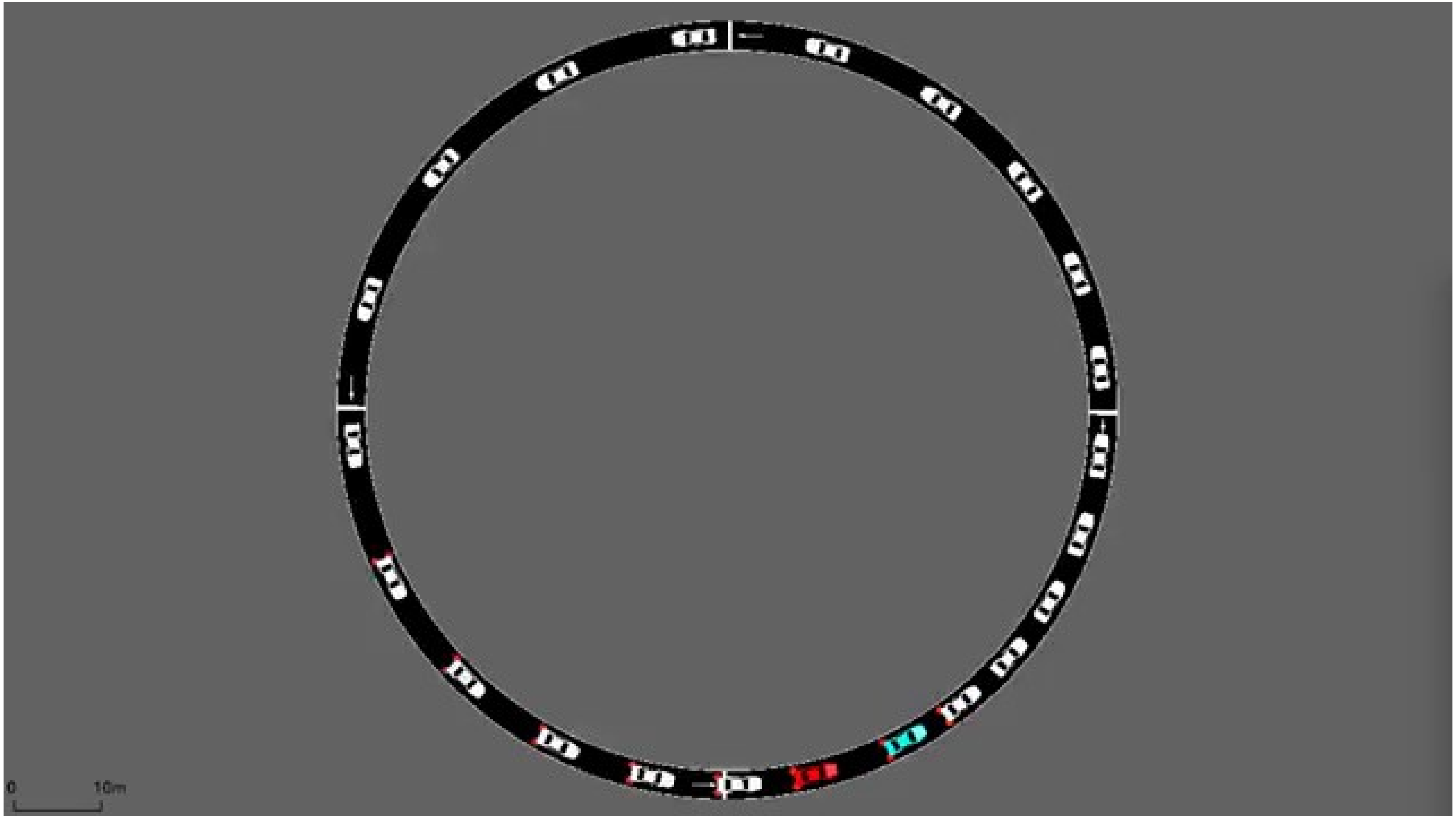
Perception Uncertainty



SCALABLE MULTIAGENT DRIVING POLICIES FOR REDUCING TRAFFIC CONGESTION (AAMAS 2021)

Jiaxun Cui, William Macke, Aastha Goyal, Harel Yedidsion, Daniel Urieli, Peter Stone

Learning Agent Research Group
The University of Texas at Austin
General Motors
Sony AI



E. Vinitsky, A. Kreidieh, L. Le Flem, N. Kheterpal, K. Jang, C. Wu, F. Wu, R. Liaw, E. Liang, and A. M. Bayen. Benchmarks for reinforcement learning in mixed-autonomy traffic. In Conference on Robot Learning, pages 399–409, 2018.

Problem Setting

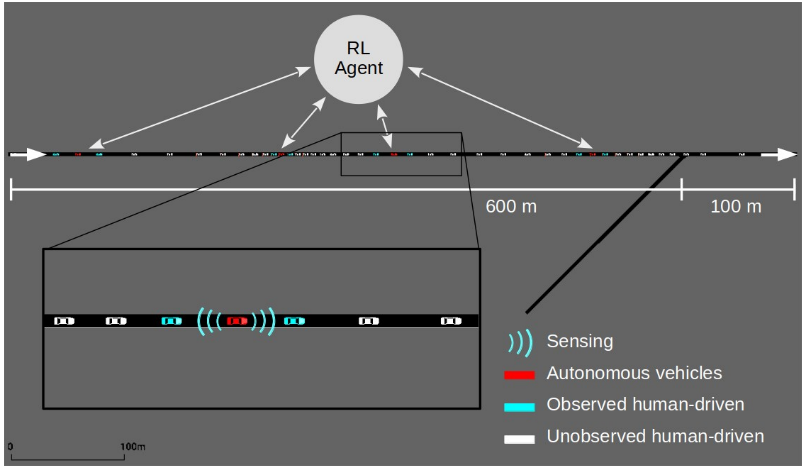
Develop a multiagent driving policy for Autonomous Vehicles(AV) in a mixed autonomy setting, and in **large-scale open** road networks

- Two-lane Merging Scenario
- Uniform Inflow
- 10% AVs and 90% Human-Driven
- Uniform AV distribution

Traffic Network: Open & Large

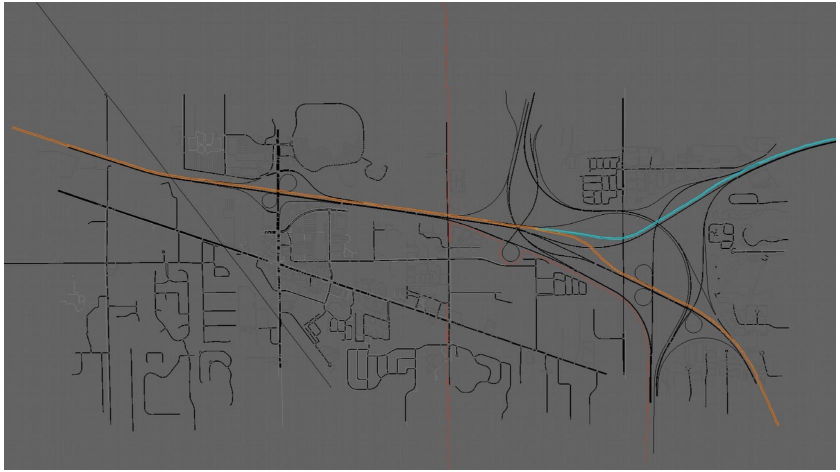
Open Network

- Short in length
- Fewer vehicles



Large Network

- Longer in length
- More vehicles



Our Solution

- For **open** and **large** **merge** network, we propose using **outflow** as an evaluation metric given a **fixed traffic inflow distribution**
- **Modular transfer** a policy trained under the small network to the segment with a similar road structure in the large network
    - Centralized RL agent[1] but the learning and policy execution only happens in the road window

SCALABLE MULTIAGENT DRIVING POLICIES FOR REDUCING TRAFFIC CONGESTION (AAMAS 2021)

Jiaxun Cui, William Macke, Aastha Goyal, Harel Yedidsion, Daniel Urieli, Peter Stone

Learning Agent Research Group
The University of Texas at Austin
General Motors
Sony AI

Experiment Result: Simple Merge

Simple Merge is an **Open** and **Small Merging** Network

With a fixed inflow rate and number of controllable autonomous vehicles

We obtain best outflow by using a time-step outflow as a reward

Table 1: Statistics of Reward Functions on Simple Merge

Reward	Average Outflow (vehs/hr)	Average Inflow (vehs/hr)	Average Speed(m/s)
Human	1559.88±2.758	1726.68±2.611	7.27± 0.029
Original Flow Reward	1690.70±6.131	1746.76±6.339	15.80±0.102
Average Speed Reward	1521.72±3.067	1560.42±4.136	**18.67**±0.106
Outflow Reward	**1801.80**±7.362	**1862.28**±7.181	15.96±0.092

The results are obtained from 100 independent evaluations and we report the mean values of metric readings accompanied with their 95% confidence interval bounds.

Reinforcement Learning for Optimization of COVID-19 Mitigation Policies

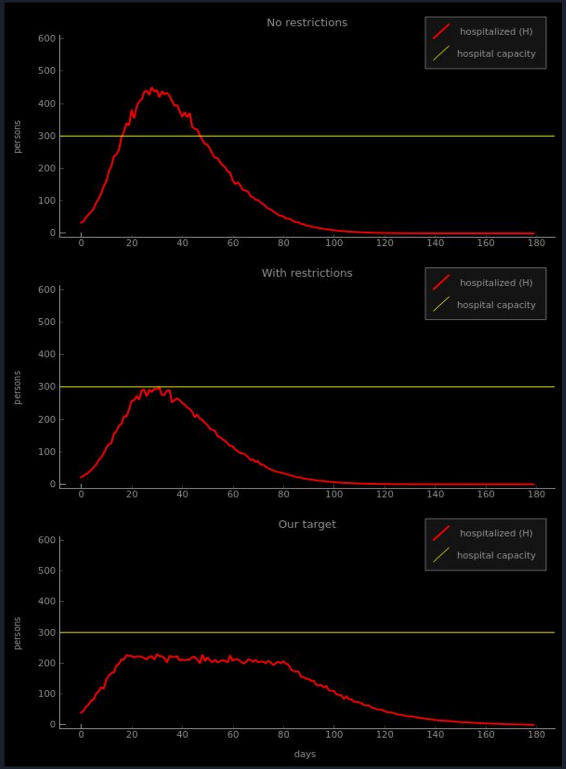
Varun Kompella*[1], Roberto Capobianco*[1, 2],
Stacy Jong[3], Jonathan Browne[3], Spencer Fox[3],
Lauren Meyers[3], Peter Wurman[1], Peter Stone[1, 3]

[1] Sony AI
[2] Sapienza University of Rome
[3] The University of Texas at Austin

*Joint First Authors, varun.kompella@sony.com, roberto.capobianco@sony.com
Paper: https://arxiv.org/abs/2010.10560
Code Repo: https://github.com/SonyAI/PandemicSimulator

Motivation

- **Goals:**
  - manage the impact of COVID-19
  - explore sequential strategies/policies to impose and relax restrictions that also favor economy

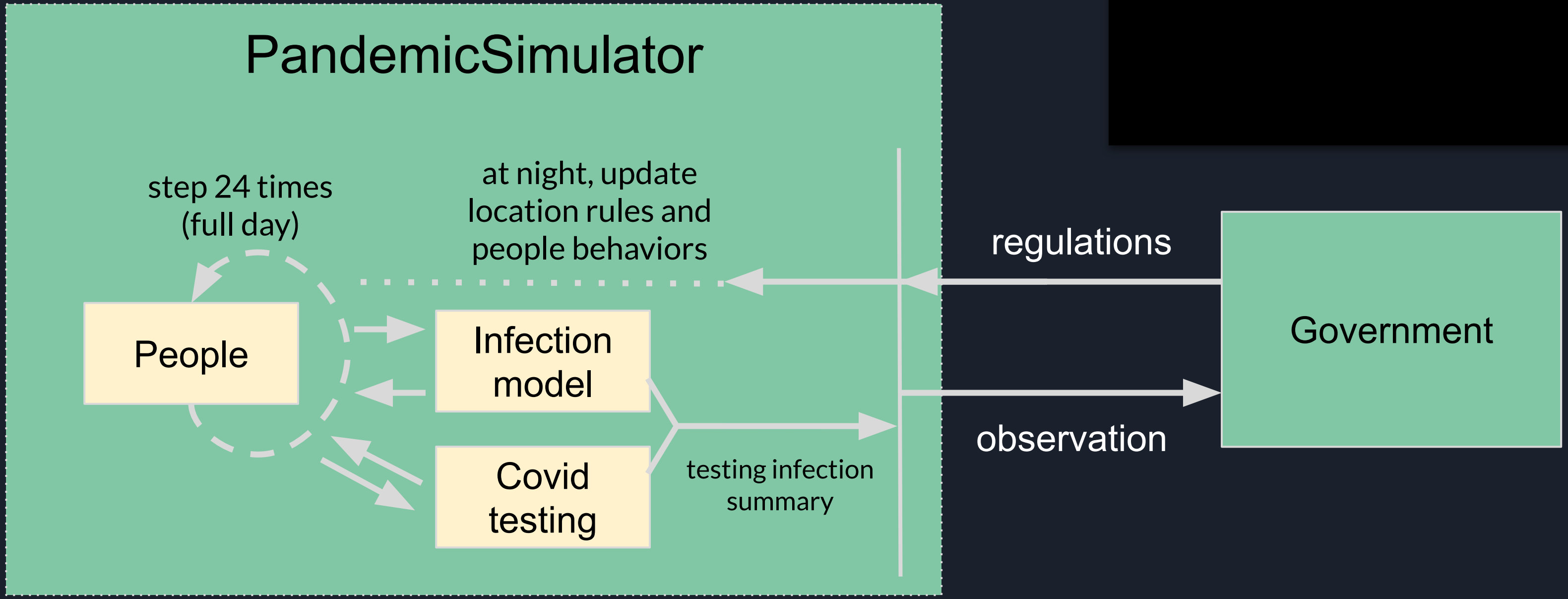


Contributions

- PandemicSimulator
    - An open-source[1] agent-based simulator that models community interactions
    - Spread of the disease is modeled as an emergent property of people's behavior
    - Models realistic effects of imperfect testing, variable spread rates among infected, flouting, contact tracing, etc.
    - OpenAI Gym interface to enable support for Reinforcement Learning (RL) libraries
- Optimize and analyse a reopening policy learned through RL

[1]https://github.com/SonyAI/PandemicSimulator

Control Loop of the Simulator



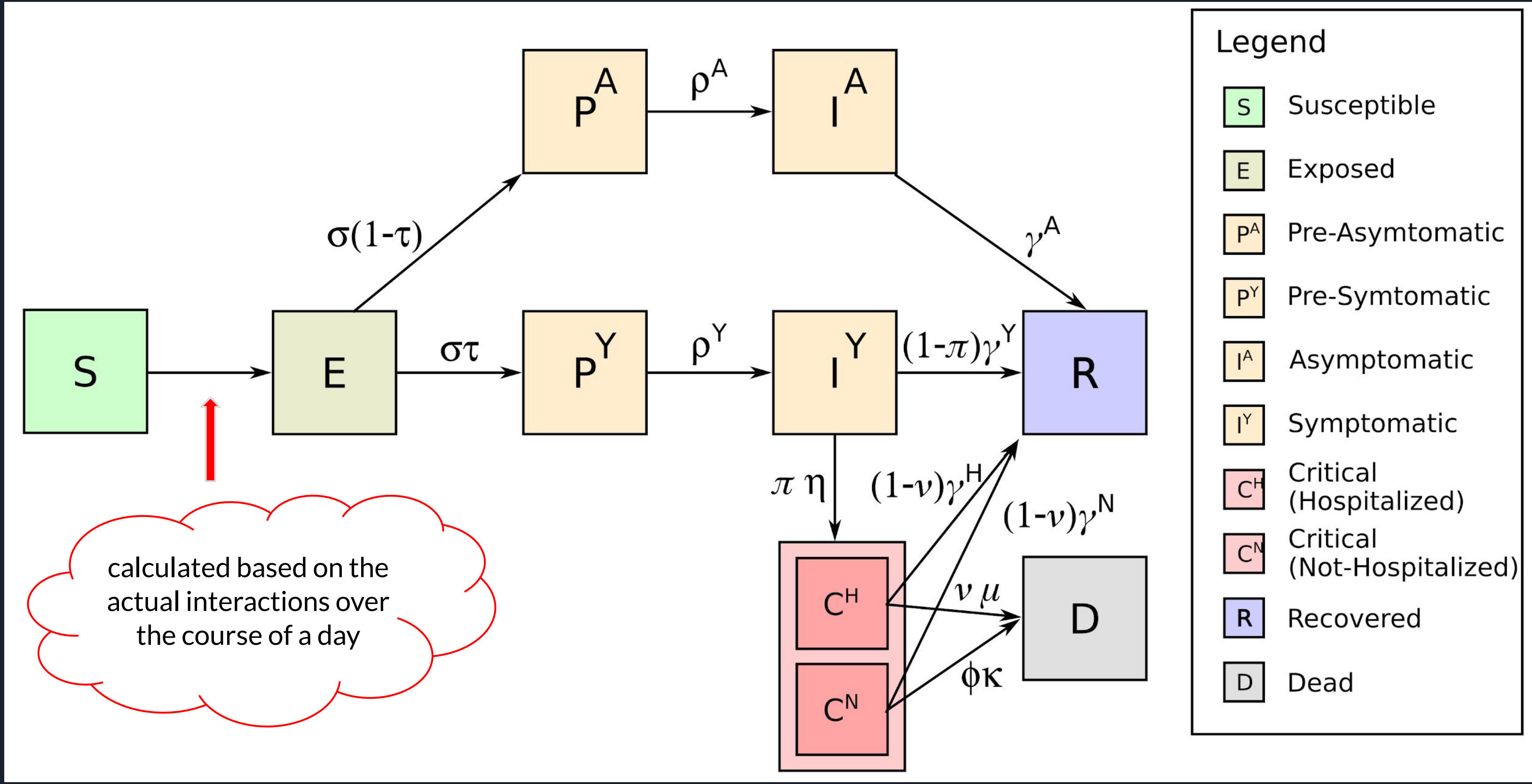
Examples of Location Rules

- Grocery Store, Office, School, Retail Store, Hair Salon
  - opening and closing hours
  - locked or unlocked
- Hospital
  - open at all the times
- These rules can be modified based on the Government decisions at any time

Examples of Person Activity Simulation

- Stochastic behaviors
  - A working person goes to an assigned office during the day
  - A child goes to an assigned school during the day
  - Each person visits each store once per week, and hair salon once per month in assigned time slots
  - At night, each person stays home or goes to a social-event (house party) twice a month
  - A person to-be-hospitalized goes to a hospital, unless when the hospital is full (in this case stays home)
  - Some people flout regulations

Infection Model



Infection Probability (S -> E)

- Infection probability for each person is calculated based on the actual contacts between people in the simulator over the course of each day
- Each person has an infection spread rate that is sampled from a bounded gaussian distribution
  - For example - super spreaders have higher rates
- Incubation period of ~2.5 (probabilistic) days before becoming infectious (and testing positive)

\

Government Actions and Observations

- Discrete Stage parameters:
  - Lock a location
  - Practice good hygiene
  - Stay at home when sick
  - Wear masks
  - Social distancing
  - Quarantine
  - Max gathering size for high/low risk persons
- Observations:
  - Infection summary (critic)
  - Testing summary (actor)
  - Stage

COVID-19: Risk-Based Guidelines

	Practice Good Hygiene / Stay Home If Sick / Avoid Sick People	Maintain Social Distancing	Wear Facial Coverings	Higher Risk Individuals (Age over 65, diabetes, high blood pressure, heart, lung and kidney disease, immunocompromised, obesity)			Lower Risk Individuals (No substantial underlying health conditions)			Workplaces Open
				Avoid Gatherings	Avoid Non-Essential Travel	Avoid Dining/ Shopping	Avoid Gatherings	Avoid Non-Essential Travel	Avoid Dining/ Shopping	
Stage 1	●			greater than 25		except with precautions	gathering size TBD			all businesses
Stage 2	●	●	●	greater than 10		except as essential	greater than 25		except with precautions	essential and re-opened businesses
Stage 3	●	●	●	social and greater than 10	●	except as essential	social and greater than 10		except with precautions	essential and re-opened businesses
Stage 4	●	●	●	social and greater than 2	●	except as essential	social and greater than 10	●	except expanded essential businesses	expanded essential businesses
Stage 5	●	●	●	outside of household	●	except as essential	outside of household	●	except as essential	essential businesses only

Use this color-coded alert system to understand the stages of risk. This chart provides recommendations on what people should do to stay safe during the pandemic. Individual risk categories identified pertain to known risks of complication and death from COVID-19. This chart is subject to change as the situation evolves.

AustinTexas.gov/COVID19 Published: May 13, 2020 APH Austin Public Health PREVENT. PROMOTE. PROTECT.

\

Reinforcement Learning (RL) Experiments

- Small town configuration
    - 1000 persons (US population age distribution)
    - 300 homes
    - 4 grocery stores (30 max visitors, 5 workers)
    - 4 retail stores (30 max visitors, 5 workers)
    - 4 hair salons (5 max visitors, 3 workers)
    - 5 offices (no visitors, 200 workers)
    - 1 school (300 students, 40 teachers)
    - 1 hospital (10 patients, 5 doctors)
    - 1 cemetery

- Rewards (costs)
    - Control infection spread
        - Negative rewards if critical above max capacity
    - Resist going to higher stages (promote reopening)
        - Negative rewards proportional to higher stages
    - Shaping rewards:
        - Negative rewards for stage changes

\

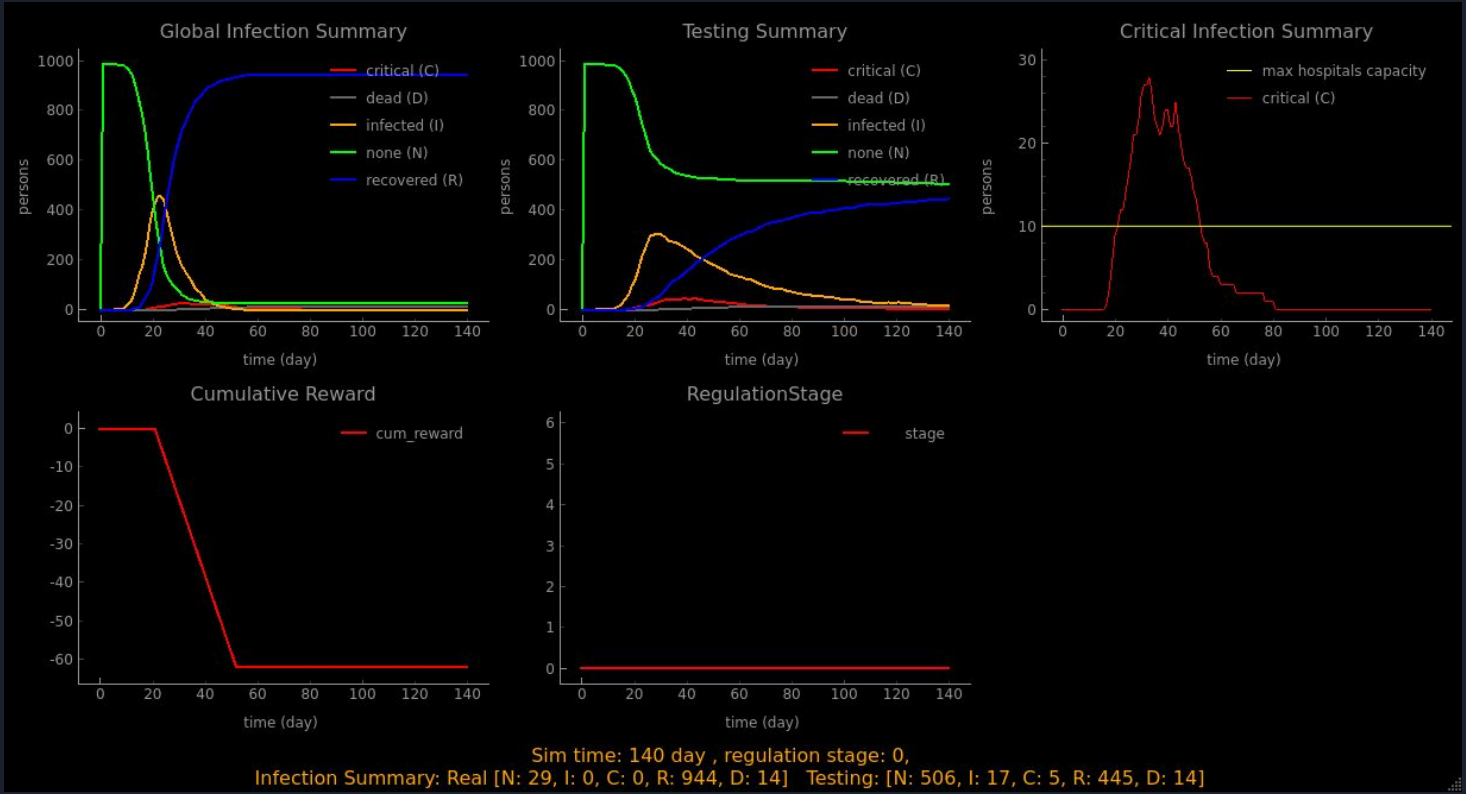
RL Training

- RL training is carried out in two parallel processes
  - Collect data
  - Update RL agent
  - Simulation speed: 0.41 secs per day, training time: ~ half hour (300k updates, 5000 days of data)

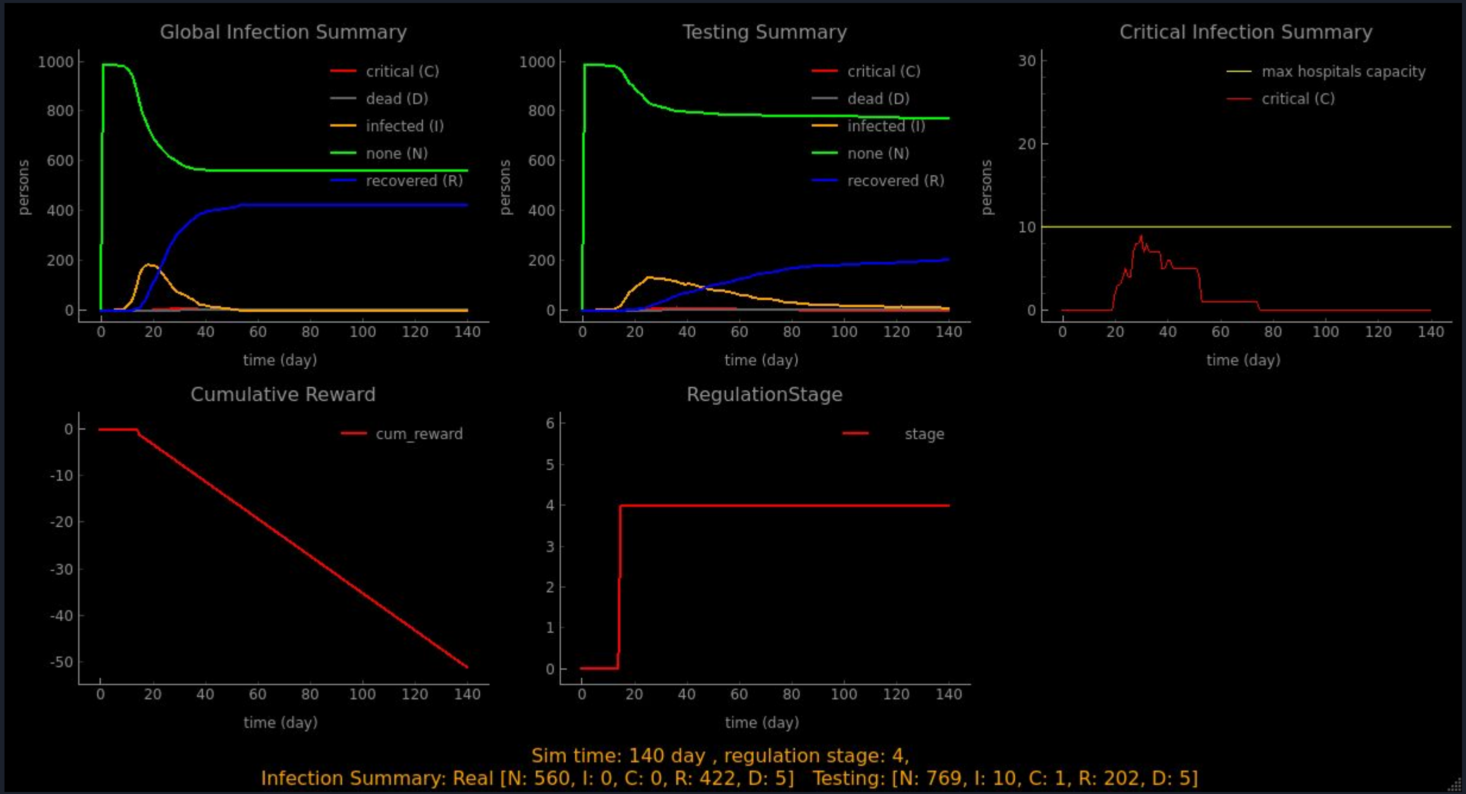
- Collect data
  - Repeat:
    - Sample a regulation (stage) from the policy
    - Iterate simulator for 24 steps
    - Get observation from the simulator
    - Compute reward
    - Add (observation, regulation, reward) to a data buffer

- Update RL agent
  - Sample a batch from the data buffer
  - Update policy and critic parameters

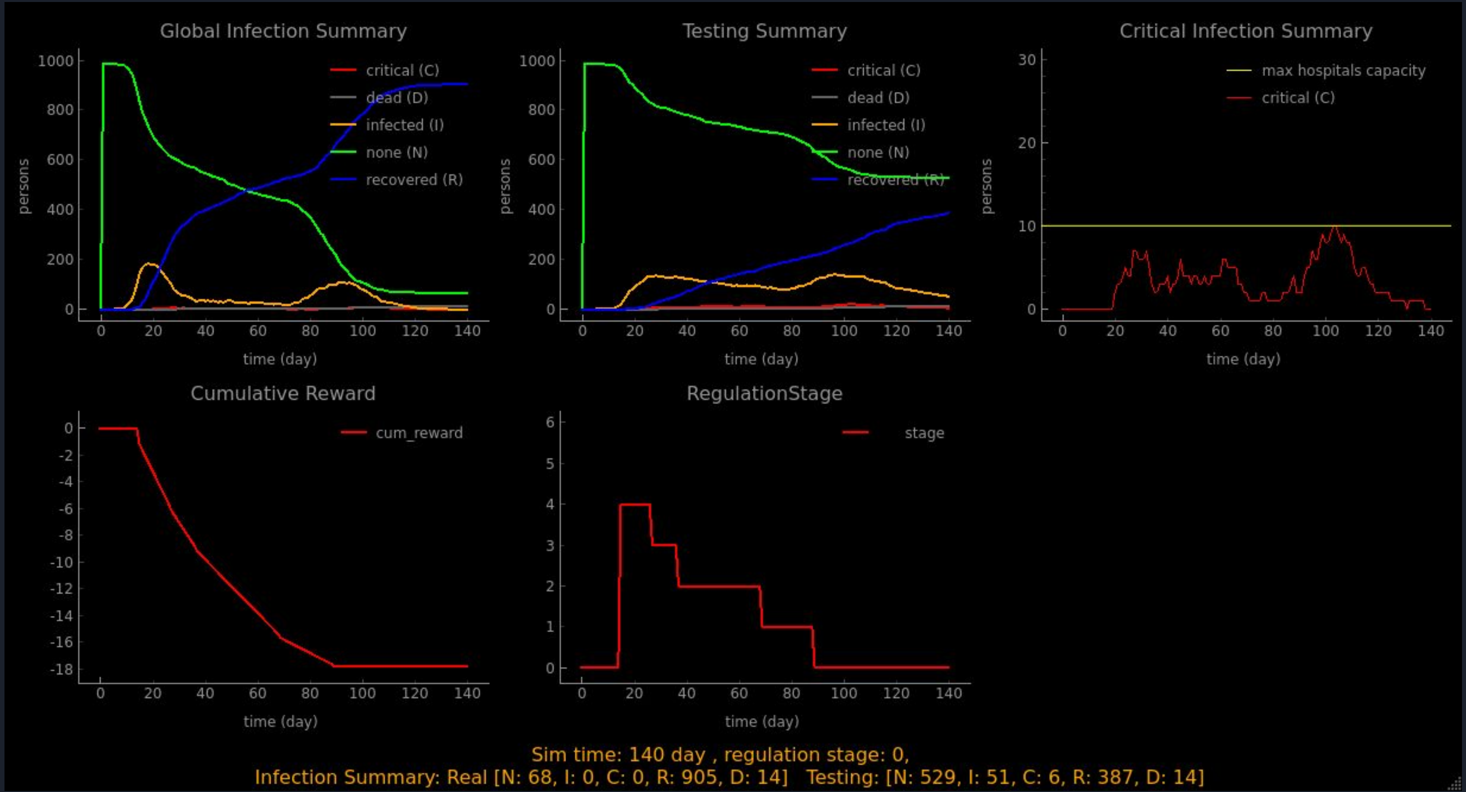
Baseline Comparison (Stage-0 Policy)



Baseline Comparison (Stage-4 Policy)

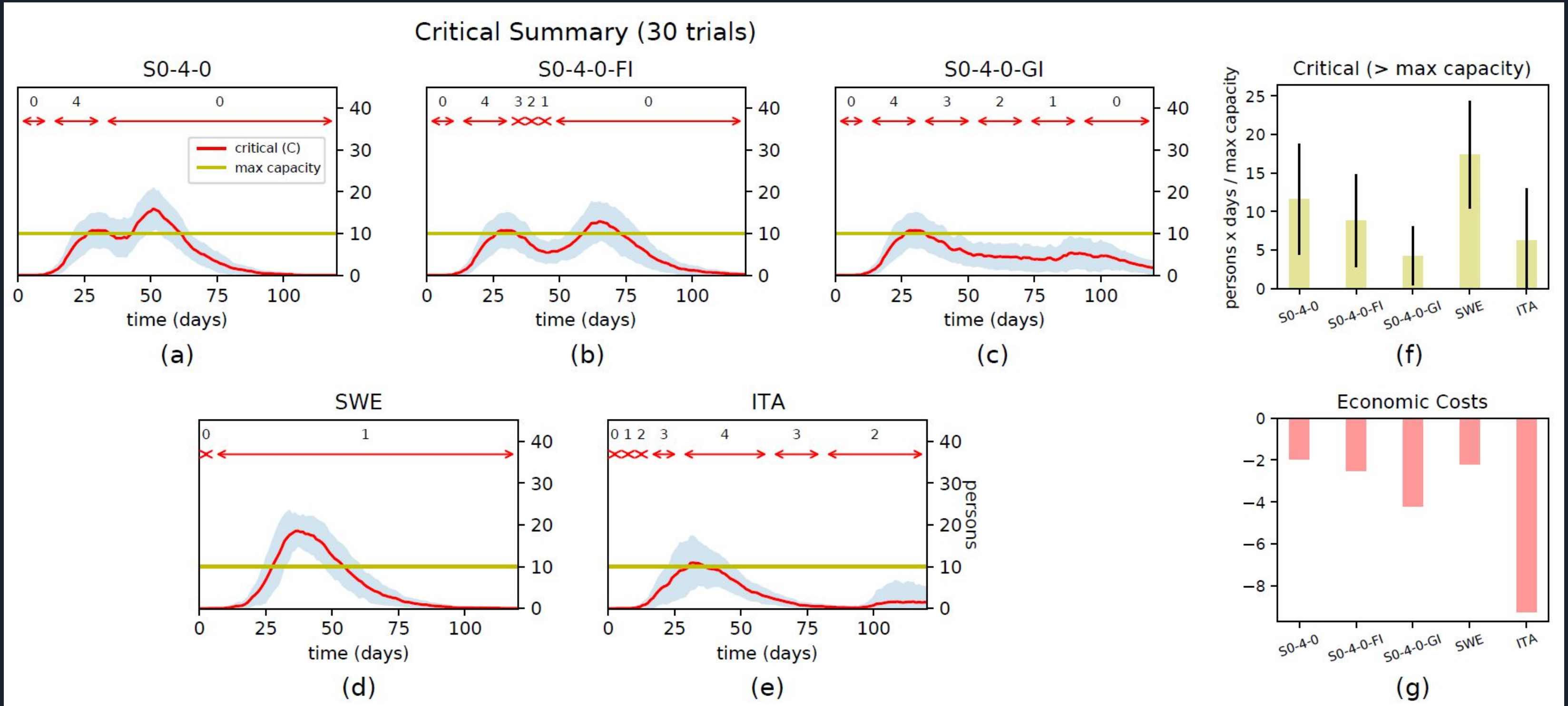


Learned Stochastic Policy



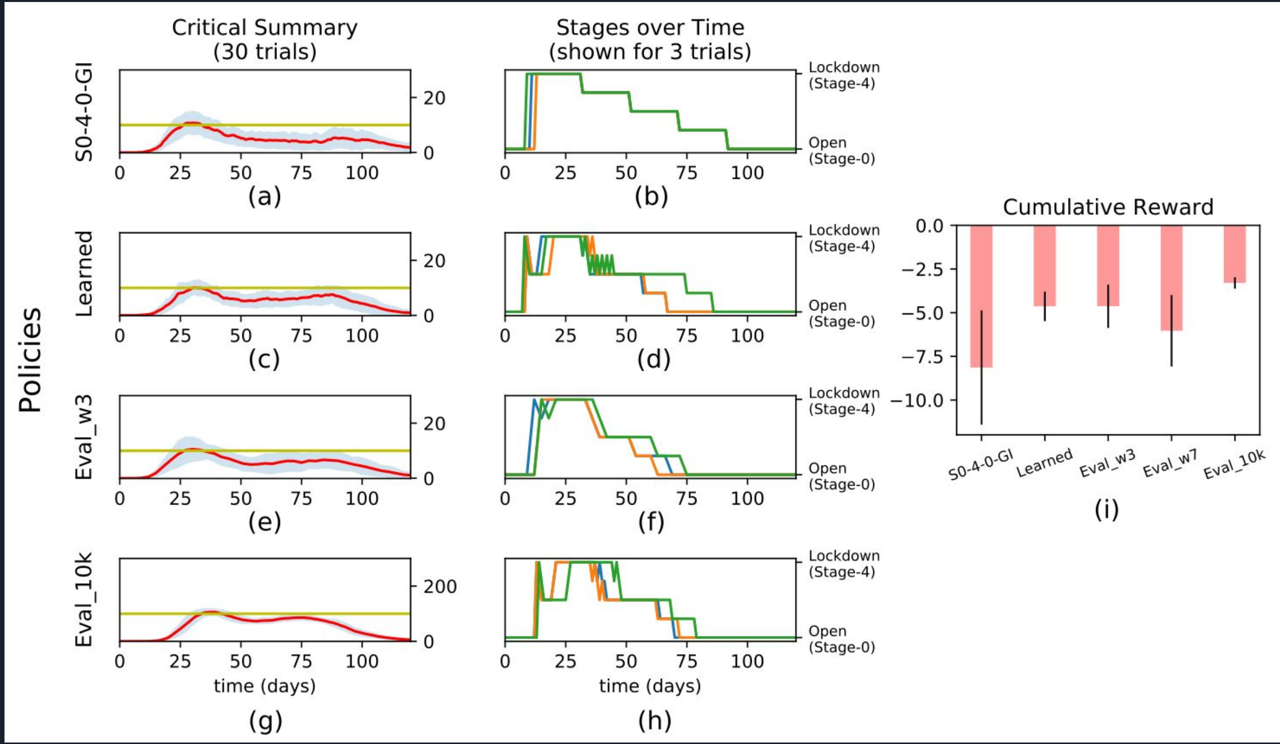
Analysis of Benchmark Policies

Regulations: 5 staged escalating restrictions (0 - 4)



Optimizing Reopening using RL

Reward = $\boxed{a\,(n^{critical} - \text{max-capacity})}$ + $\boxed{b\ stage^{c}}$



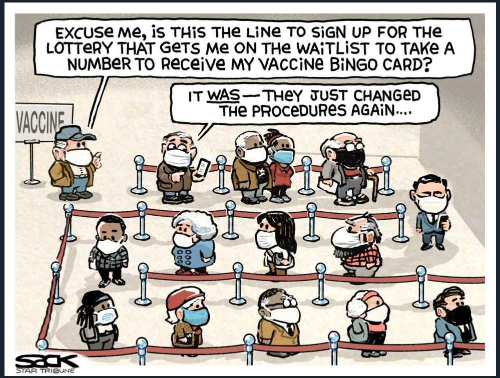
Recap

- Created an open source software to simulate pandemics in a “sim-city” like environment
  - https://github.com/SonyAI/PandemicSimulator

- We calibrated our simulator using real-data, did sensitivity analysis, added contact tracing, testing, etc.

What if there is vaccination in the horizon?

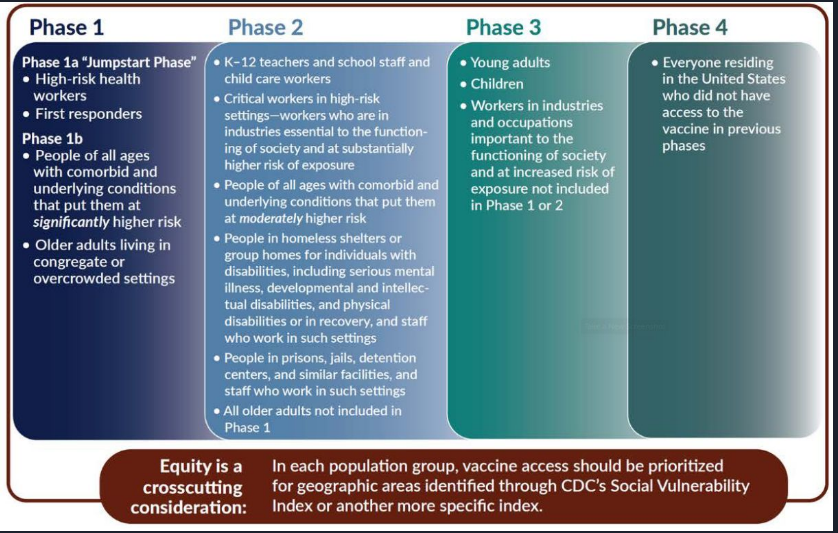
Vaccination framework:

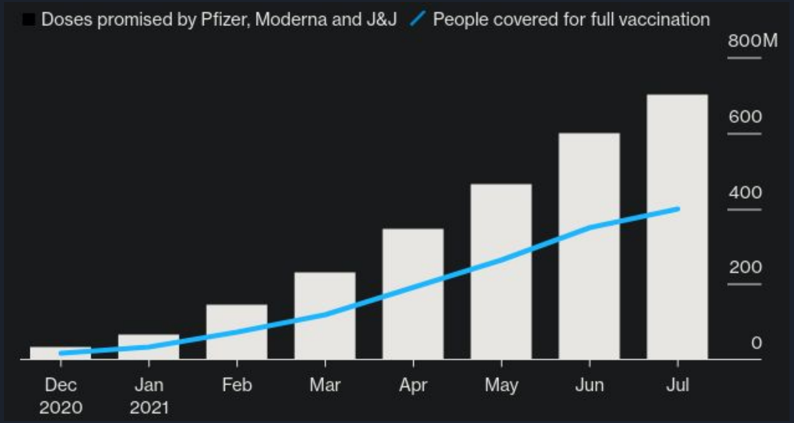


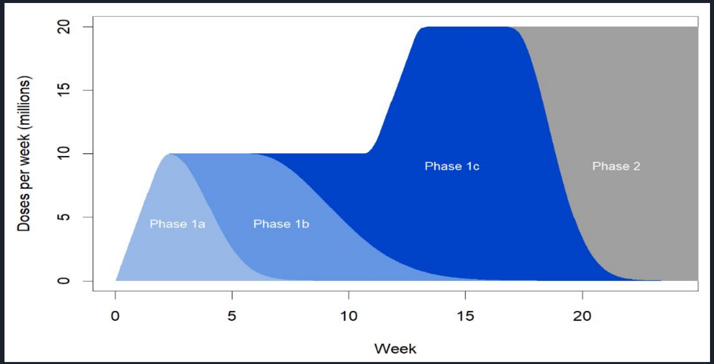
- Vaccination Centers
  - A person visits the center to get a vaccination
  - Maintains local vaccination summaries
- Person Routines
  - Get on the queue to get a shot at one of vaccination centers
- CDC
  - Controls supply, eligibility, phases, etc.
  - Vaccine allocation model
    - Specifications for vaccines, rollout phases, availability chart, etc.
- Added state Information
  - Vaccination state for each person
  - Global vaccination summary
  - Vaccination start date and supply rate

Vaccine allocation model

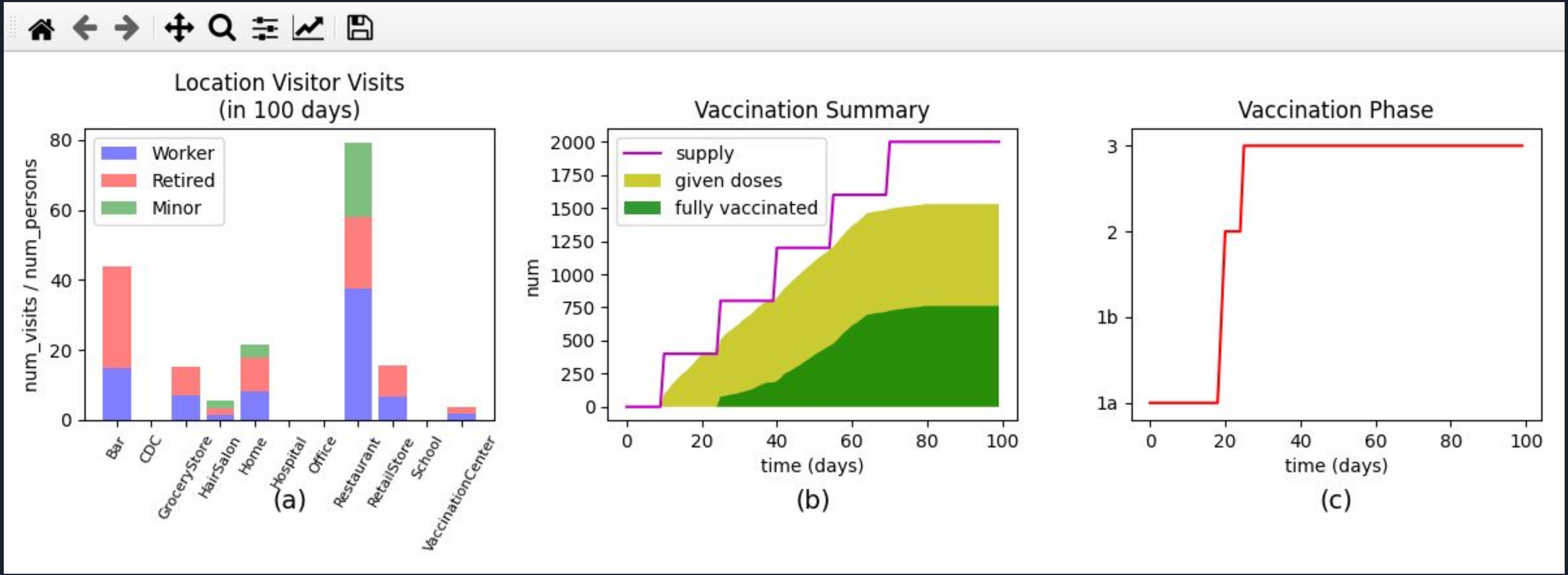
References:
[1] www.nap.edu/resource/25917/FIGURE%20-%20A%20Phased%20Approach%20to%20Vaccine%20Allocation%20for%20COVID-19.pdf
[2] www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-12/slides-12-20/02-COVID-Dooling-508.pdf
[3] www.bloomberg.com/news/articles/2021-02-18/how-many-vaccine-doses-are-available-u-s-should-see-a-surge





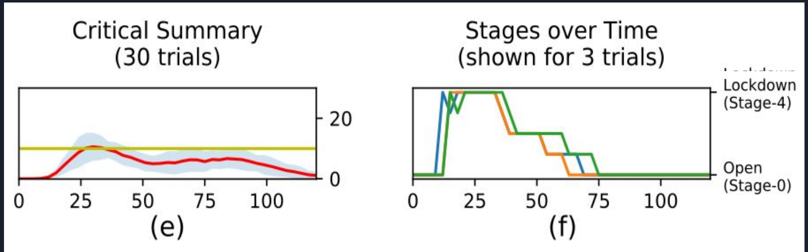


1000 person simulation run (preliminary result)
(vaccination_start_day = 10, supply_interval=15 days)



Summary

- Learned an RL policy that optimizes a reopening strategy balancing infection spread and economic costs (AAAI Symp, submitted to JAIR).
  - Main insight: The best strategy is to switch gradually from complete lockdown to no-restrictions. It is very expensive (economically) to stop infection spread entirely, so longer lockdowns are sub-optimal.



- Also: “Multiagent Epidemiologic Inference through Realtime Contact Tracing”
  - Thursday S6: Reinforcement Learning 4

Next steps (there are many!)

- Conduct longer experiments with larger populations
- Add more types of locations
- Higher fidelity model of schools
- Explore structured stage-policies
- Try different learning algorithms
- Post-process network results for explainability
- Finish vaccination model and run RL experiments!

Conclusions and Future Work

- Introduced an RL methodology for optimizing adaptive mitigation policies aimed at balancing economy and infection spread
- Introduced an open-source agent based simulator where pandemics can be generated through individual interactions in a community
- Future work:
    - Explore fine-grained policies
    - Test various testing/contact tracing strategies

Next steps

- Link person’s vaccination state and infection state
- Generate infection summary plots for different configurations:
  - Vaccination start date
  - Regulation stage
  - Vaccination supply rate
- Run RL!

Reinforcement Learning for Optimization of COVID-19 Mitigation Policies

Varun Kompella*[1], Roberto Capobianco*[1, 2],
Stacy Jong[3], Jonathan Browne[3], Spencer Fox[3],
Lauren Meyers[3], Peter Wurman[1], Peter Stone[1, 3]

[1] Sony AI
[2] Sapienza University of Rome
[3] The University of Texas at Austin

*Joint First Authors, varun.kompella@sony.com, roberto.capobianco@sony.com
Paper: https://arxiv.org/abs/2010.10560
Code Repo: https://github.com/SonyAI/PandemicSimulator