THE HUMAN-SWARM PROJECT

SKANDA VAIDYANATH

USC INSTITUTE FOR CREATIVE TECHNOLOGIES

(SUPERVISORS: KALLIRROI GEORGILA, DAVID TRAUM)

7/8/2019

OUTLINE

- Motivation for and description of our system
- Breaking down the system and focusing on specific modules
- The Negotiation Module
- Reinforcement Learning
- Experiments and Results
- Future Work
- The Search Strategies Module

CALIFORNIA WILDFIRES

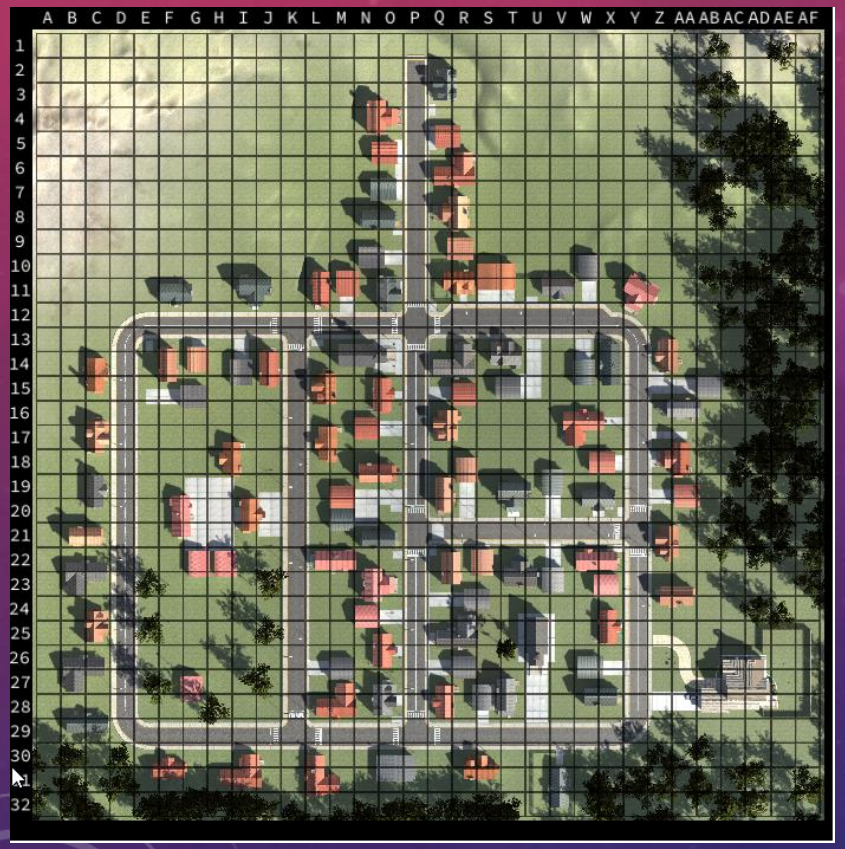
Since the start of the year, 1,349 **wildfires** in the state have burned 12,247 acres of land, according to **California** Department of Forestry and Fire Protection, also known as CalFire. Of those **fires**, 178 have been on federal land. There have been 236 new **wildfires fires** in the past week alone. Jun 13, 2019

Deadly wildfires in Northern California

Firefighters walk along State Route 20, one of two roads closed by a fast-moving wildfire near Clear Lake in California on Sunday. The fire has burned about 17,000 acres and destroyed more than 20 homes. About 12,000 people have fled the fire or been advised to evacuate. Farther north, another blaze killed a firefighter. Story, A6.

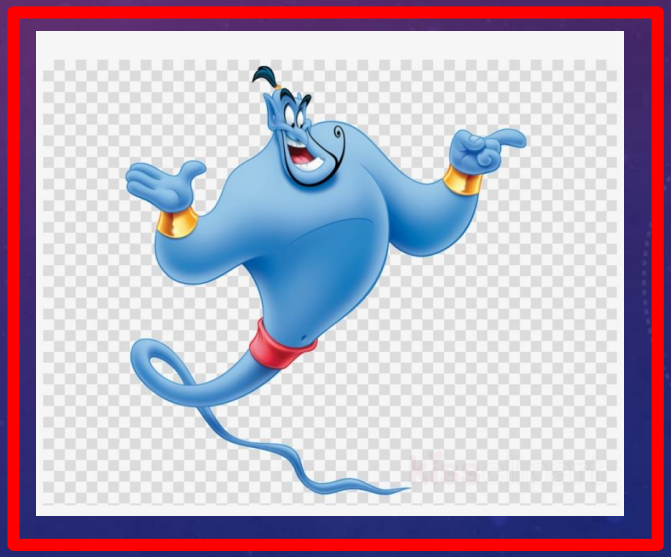
WILDFIRE RESCUE OPERATIONS

OUR SYSTEM



The drones

The Operator



The Spokesperson

THE BIG PICTURE

- The Operator and the Spokesperson control the drones i.e. call out commands/plays that eventually help save the civilians in the town. We control the drones similar to how we would control players in a sports game, for example, setting fields in cricket!
- They could ask drones to watch the fire progression, search a particular area, engage in negotiation with a civilian group, guide them to safety, etc.
- This town has three different civilian groups at random locations on the map. The big picture is to find these civilian groups (search for them), engage (negotiate) with them and convince them to leave their homes and guide them to safety.

THE MODULES WE FOCUS ON

- The Negotiation Module
- The Search Strategies Module

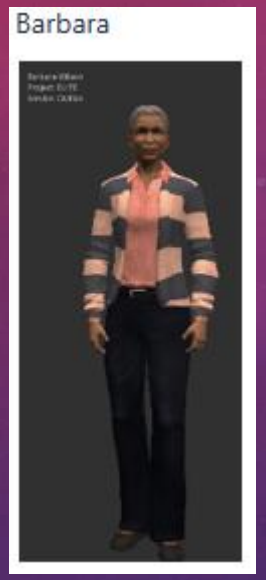
THE NEGOTIATION MODULE

- The goal is to convince the civilian group to leave the house so that the drone can guide the group to safety.
- There are three different types of civilian groups in our setup – the babysitter, the old couple and the stubborn couple. Each of them require different strategies to be convinced and different ways to be guided to safety.
- The Spokesperson issues commands to the drone to warn the group, to open a channel for the Spokesperson to negotiate with them directly, to interrupt the Operator or guide them to safety.
- Note that we do not know in advance which of the three groups the drone is currently engaged with, forcing us to come up with a strategy that works well for all three groups.
- Factors that could affect our negotiation strategy are how busy the Operator is, how much time we have before the fire reaches the group, etc.

THE BABYSITTER AND THE CHILD

Usually convinced with a warning (or by the Spokesperson or by the Operator). Need the drone to guide them out.

THE OLD COUPLES



Usually convinced by the Spokesperson (or the Operator). Need to be guided out in a vehicle.

THE STUBBORN COUPLE

Ali

Courtney

Can only be convinced by the Operator. Happy to guide themselves out.

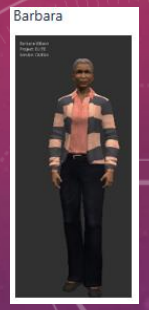
NEGOTIATION SCENARIOS

WARN

The fire is fast approaching! Please evacuate your homes!

NO! Go away!

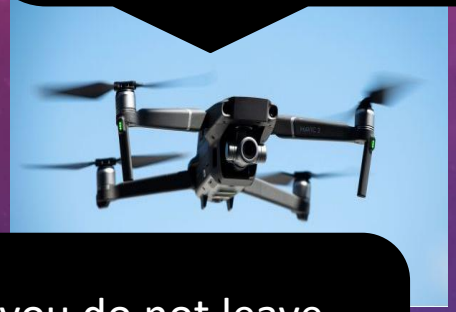
Okay! Could you please help us out?!



NEGOTIATION SCENARIOS

LET ME NEGOTIATE WITH THEM

This is the Spokesperson speaking! Please leave your home so that you are not hurt in the fire!

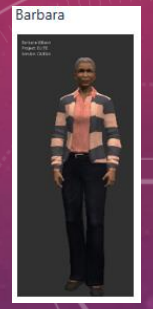


If you do not leave, you will be burned down with your house!

But we do not want to leave. It is not safe out there.

Okay we will leave!

Okay! Could you please help us out?!



NEGOTIATION SCENARIOS

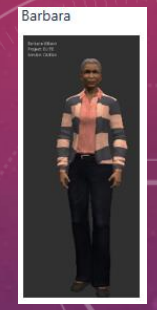
INTERRUPT THE OPERATOR

This is the Operator speaking! Please evacuate your homes immediately!

I'm afraid there is no time for that now! Please leave now. The drone will guide you out.

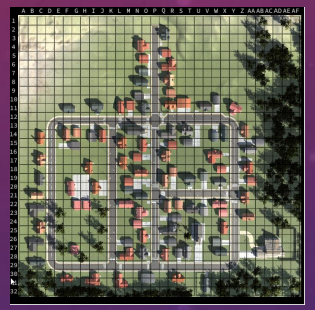
But we need to water our garden first!

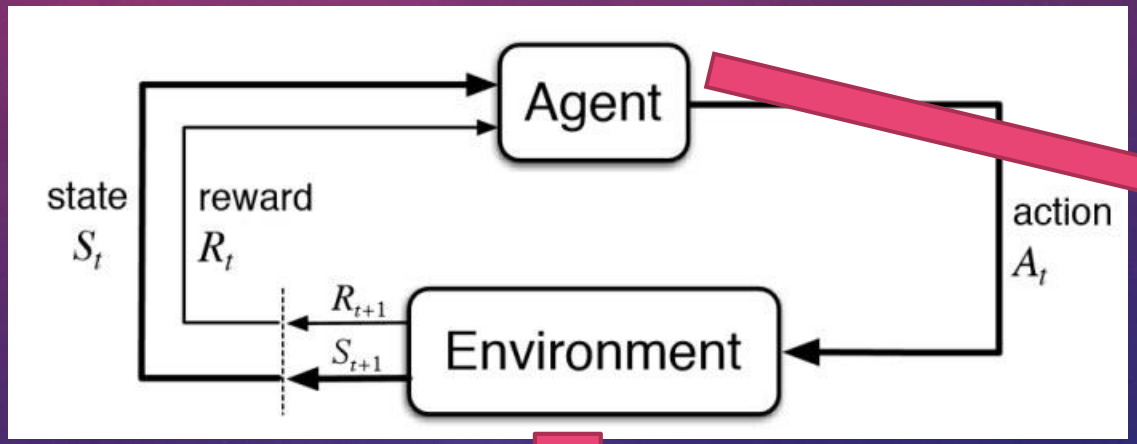
Okay fair enough. We'll guide ourselves out.



REINFORCEMENT LEARNING

Reinforcement learning (**RL**) is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

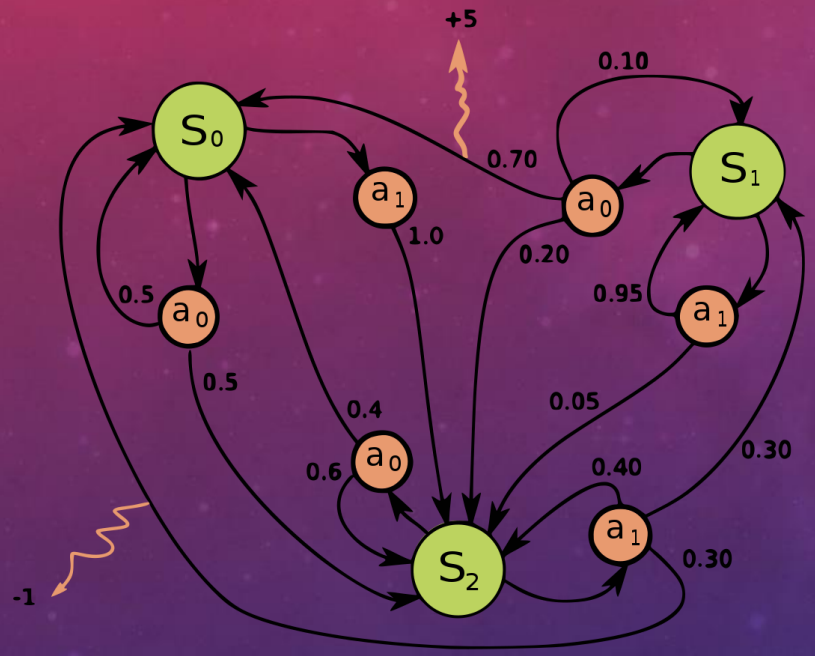


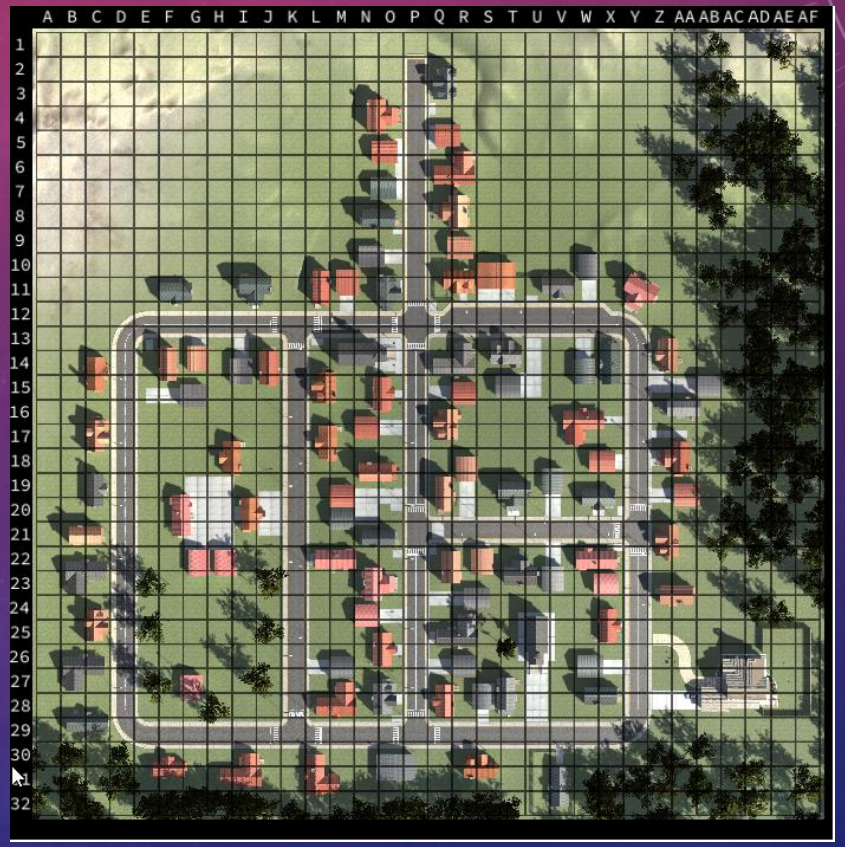


WHY USE RL HERE?

- We want to come up with a ‘policy’ that can accurately determine what our next step in the negotiation should be so that we can eventually convince the group to leave and use the appropriate guiding strategy to save them.
- For example, interrupting the Operator may be the most effective strategy in terms of convincing the groups, but the Operator could be busy or we may not have the time to interrupt the Operator all the time. Maybe we only have the time to issue a warning.
- All this becomes much harder since we do not know which group we are talking to.

MODELLING THE PROBLEM





THE STATE SPACE

- How busy is the Operator?
- Status of the group? Is the group currently being monitored, warned, negotiated with, ready to move, etc.
- How long before the fire reaches us? We call this dead time.
- What guide strategy does the group prefer?
- Status/ history of the negotiation.

This gives us 1440 states.

THE ACTION SPACE

- 0) Wait
- 1) Interrupt Operator
- 2) Query for Guide preference
- 3) Drone guide
- 4) Call for vehicle
- 5) Warn
- 6) Spokesperson negotiation

THE REWARD FUNCTION

- +5000 is the group is saved i.e. convinced and appropriately guided
- -5000 if the group is killed.
- -300, -800, -1300, -1800 for interrupting the Operator
- Reward shaping to make sure that agent does not play any ‘bad moves’ (Eg: Calling a guide when the group is not ready to move.)

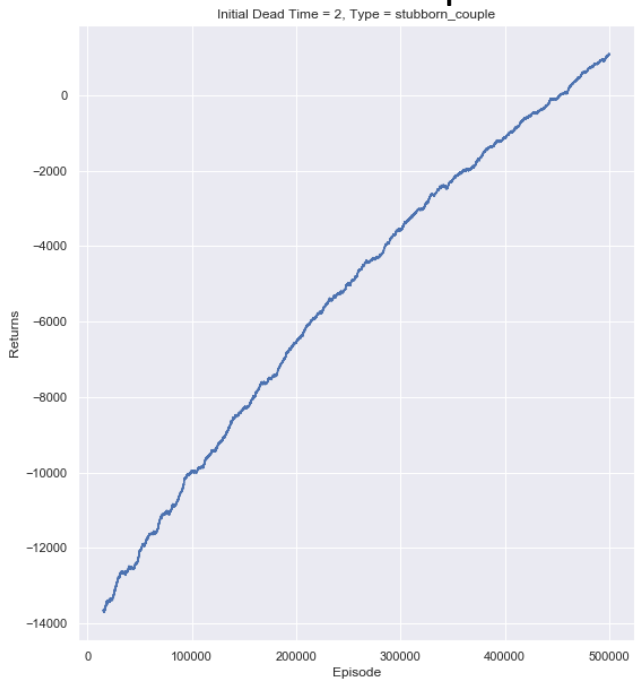
RL ALGORITHMS USED

- Monte Carlo
- Off Policy Monte Carlo with Weighted Importance Sampling
- Q-Learning
- SARSA
- Expected SARSA

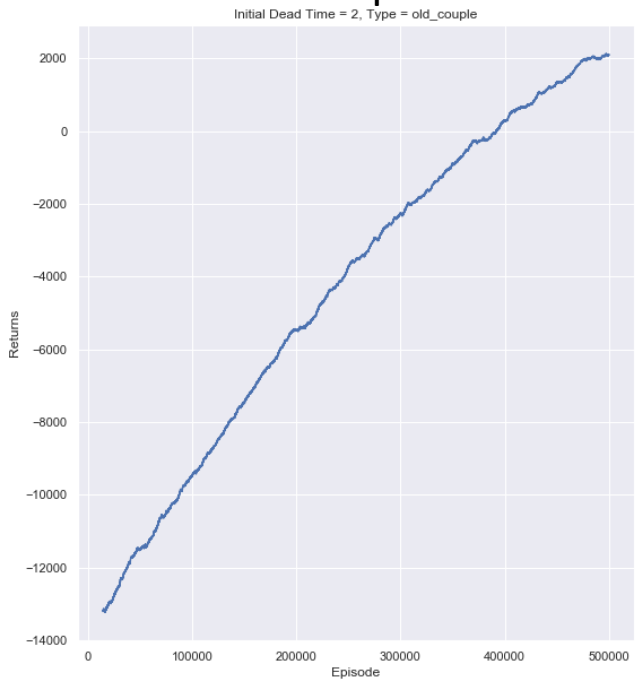
TRAINING CURVES INITIALIZED TO DEAD TIME 2

X axis: Episodes, Y axis: Returns

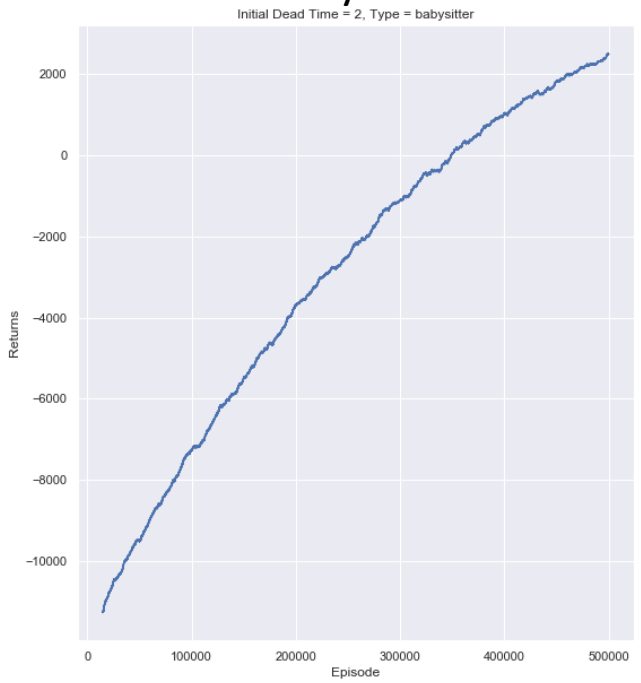
Stubborn Couple



Old Couple



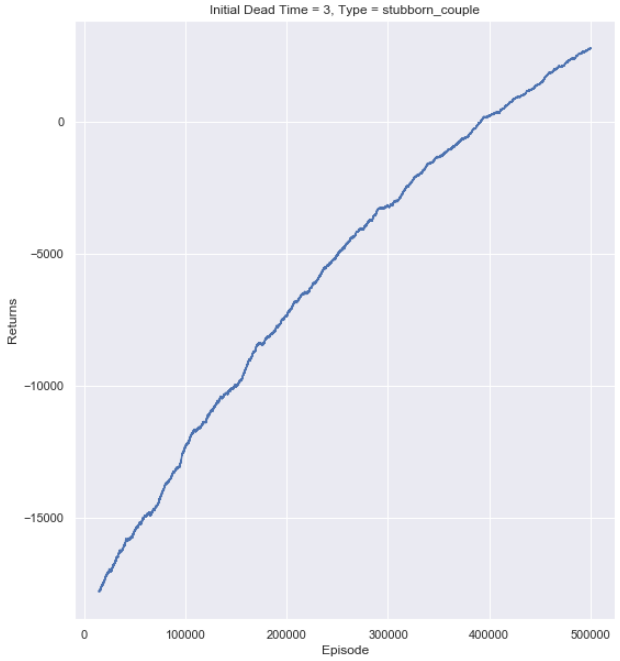
Babysitter



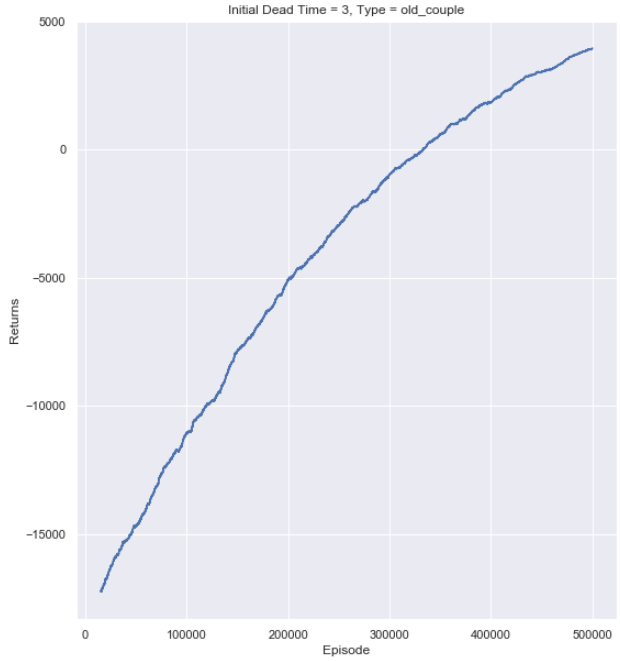
TRAINING CURVES INITIALIZED TO DEAD TIME 3

X axis: Episodes, Y axis: Returns

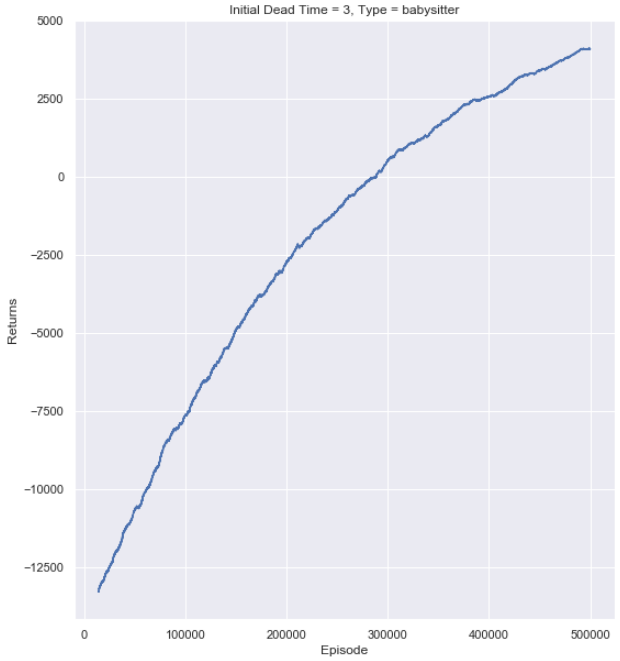
Stubborn Couple



Old Couple



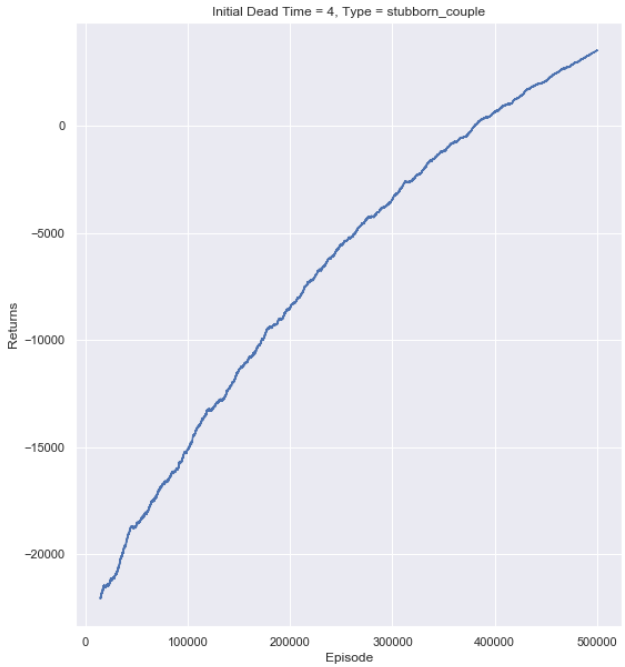
Babysitter



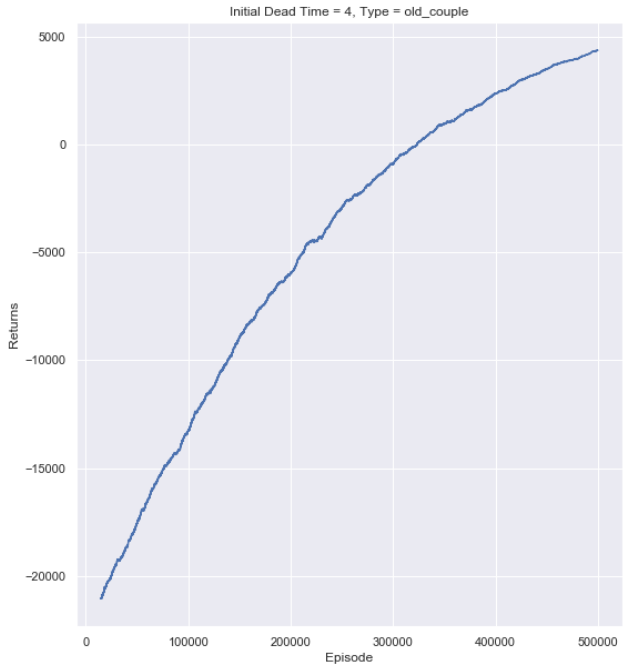
TRAINING CURVES INITIALIZED TO DEAD TIME 4

X axis: Episodes, Y axis: Returns

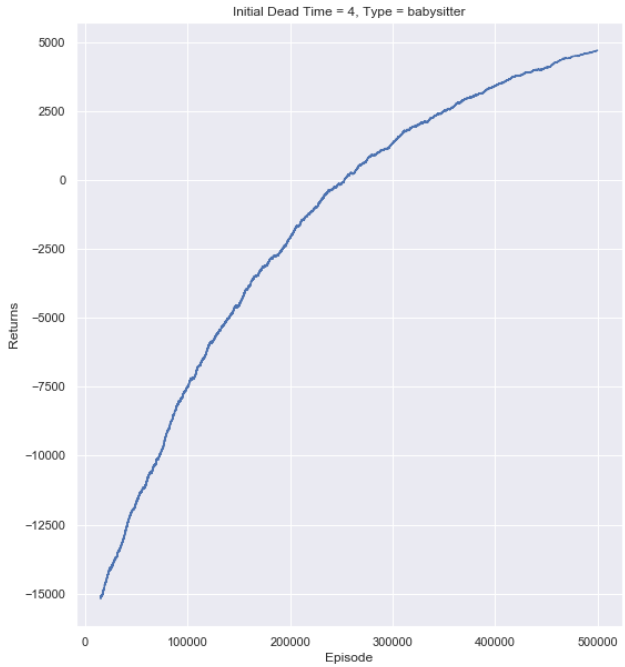
Stubborn Couple



Old Couple

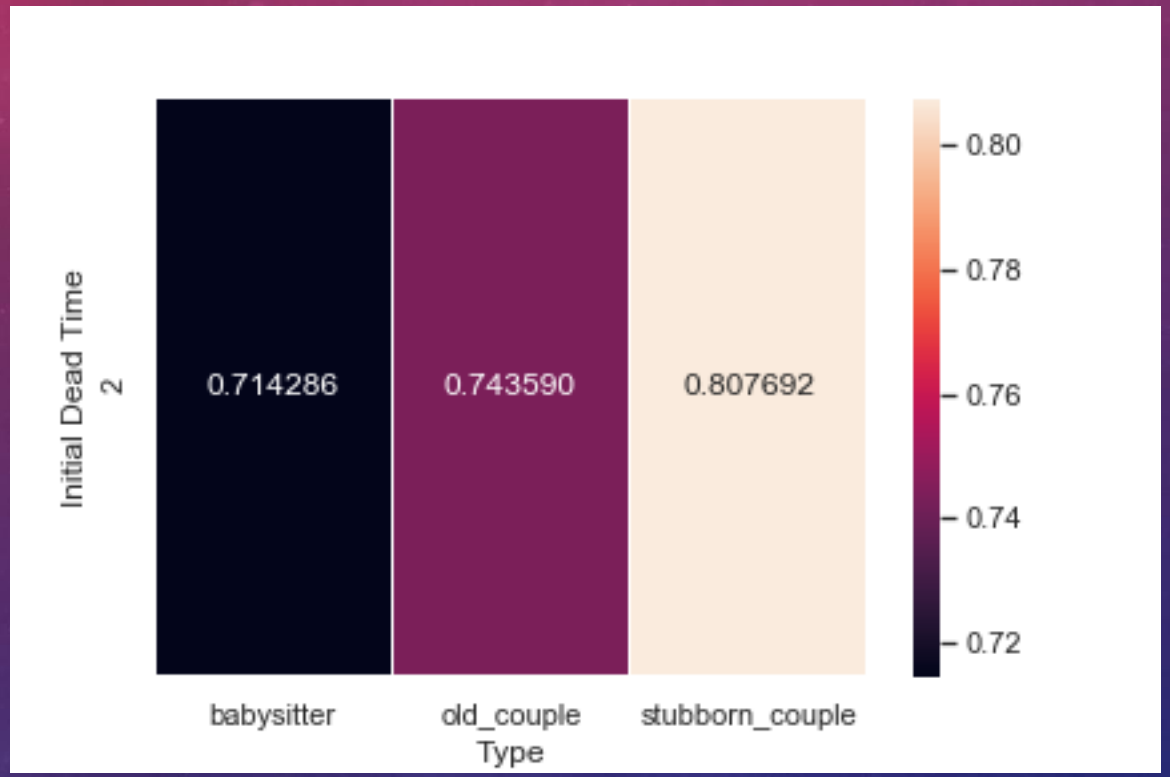


Babysitter



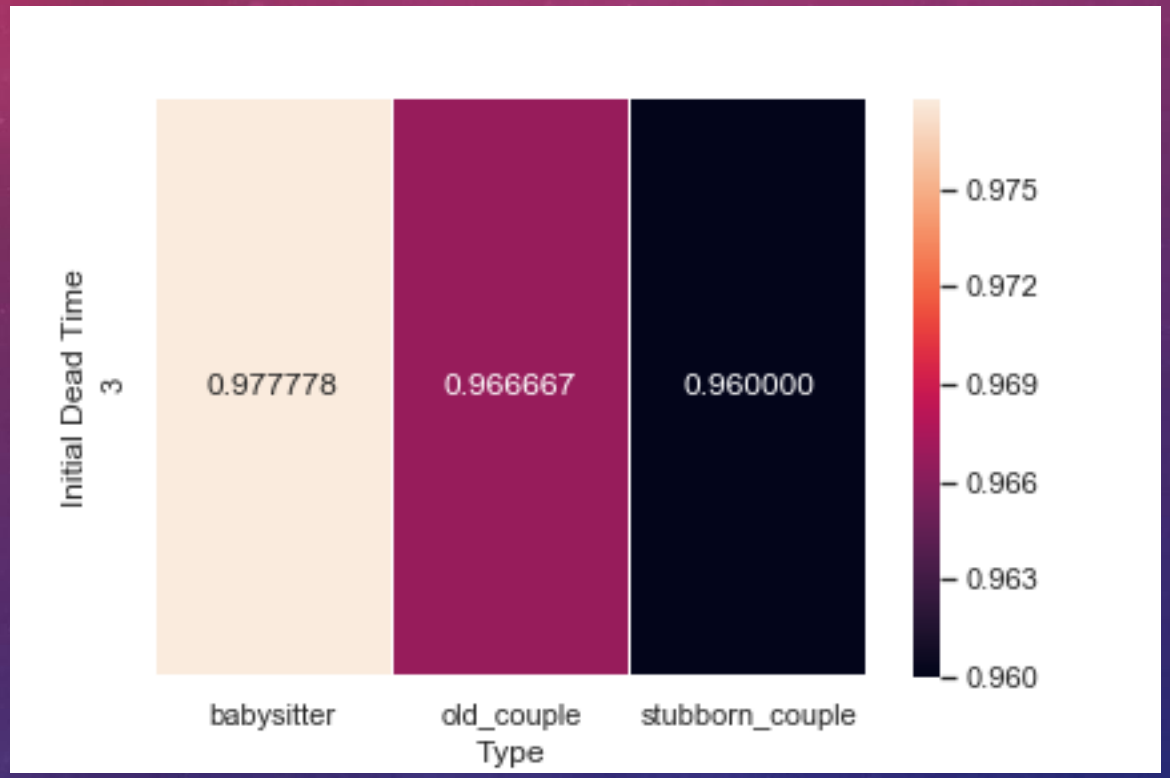
TESTING STATISTICS INITIALIZED TO DEAD TIME 2

Fraction of people saved under each initialization



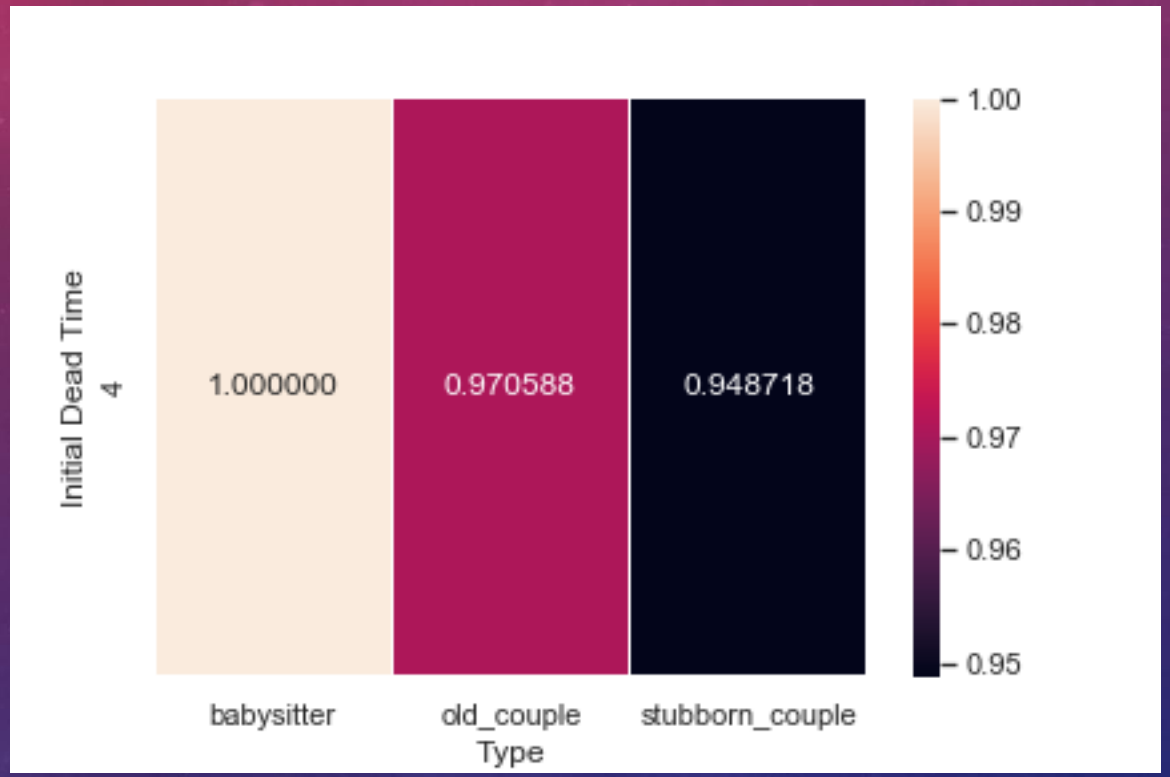
TESTING STATISTICS INITIALIZED TO DEAD TIME 3

Fraction of people saved under each initialization



TESTING STATISTICS INITIALIZED TO DEAD TIME 4

Fraction of people saved under each initialization



LOG SAMPLES

```
Episode: 25
stubborn_couple
[OP_BUSY=2, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.]
WARN
[OP_BUSY=1, STATUS=WARNING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
WAIT
[OP_BUSY=0, STATUS=WARNING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
WAIT
[OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
INTERRUPT-OPERATOR
[OP_BUSY=3, STATUS=ON, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
WAIT
[OP_BUSY=3, STATUS=ON, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
WAIT
[OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).]
QUERY-FOR-GUIDE-DETAILS
[OP_BUSY=0, STATUS=SAVED, DEAD_TIME=3, GUIDE_TYPE=SELF, NEGOTIATIONS=WARN(s).]

Episode Return is 4700
```

LOG SAMPLES

```
Episode: 1
old_couple
[OP_BUSY=3, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.]
SP-NEGOTIATE
[OP_BUSY=3, STATUS=SPN, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).]
WAIT
[OP_BUSY=3, STATUS=READY_TO_MOVE, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).]
QUERY-FOR-GUIDE-DETAILS
[OP_BUSY=2, STATUS=READY_TO_MOVE, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).]
VEHICLE GUIDE
[OP_BUSY=1, STATUS=SAVED, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).]

Episode Return is 5000
```

FUTURE WORK

- Fine-tune the policies learned for the individual initializations of the dead time. (in progress)
- Train a policy that works for any initialization of the dead time. (in progress)
- Integrate it into the final system.

SEARCH STRATEGIES MODULE

- Given a rectangular block with R rows and C columns to be searched by N drones, find the optimal coordination of these drones to search the block in the least amount of time and also making sure that not too many squares are burned before searching.
- Some squares in the block could be likely to burn down faster, depending on the location of the fire.
- One or more of the initial N drones may drop out of the search midway for whatever reason (they have been reassigned or they drop down to engage with a civilian group, etc.). In that case, the other drones need to recalibrate to search the area in the most optimal way.
- We have developed an environment for simulation for this problem as well.
- Due to the complexity of the problem, we plan on using function approximation to solve the problem.

THANK YOU!

Sources for Images: Google Images, Wikipedia, MxR Team ICT