

SIMULATION-BASED REINFORCEMENT LEARNING FOR SOCIAL DISTANCING

by

F. Ege Hoşgüngör

213150

A Dissertation

Submitted to the Department of Engineering & Informatics

University of Sussex

In Partial Fulfilment of the Requirements

For the Degree of Master of Science

June 2020

TABLE OF CONTENTS

[1 ACKNOWLEDGEMENT 2](#_Toc47556012)

[2 ABSTRACT 3](#_Toc47556013)

[3 INTRODUCTION 3](#_Toc47556014)

[4 RELATED WORK 4](#_Toc47556015)

[Flocking 4](#_Toc47556016)

[5 METHODOLOGY 5](#_Toc47556017)

[6 SOCIAL DISTANCING 5](#_Toc47556018)

[7 THEORETICAL FRAMEWORK 5](#_Toc47556019)

[8 RESEARCH DESIGN 5](#_Toc47556020)

[9 POLICY OPTIMIZATION 5](#_Toc47556021)

[10 RESULTS 5](#_Toc47556022)

[11 DISCUSSION AND CONCLUSION 5](#_Toc47556023)

[12 REFERENCES 5](#_Toc47556024)

[13 Appendix 5](#_Toc47556025)

# ACKNOWLEDGEMENT

The report may be freely copied and distributed provided the source is acknowledged.

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Prof. T. Nowotny, whose expertise was invaluable in the formulating of the research topic and methodology in particular.

In addition, I would like to thank my parents for their wise counsel and hospitality during these rough times. You are always there for me. Finally, there are my friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to rest my mind outside of my research.

# ABSTRACT

Through agent-based modelling, and standard reinforcement learning algorithms at scale, we found AI agents can give insights about ongoing epidemics by simulating the disease. An epidemic simulation has been created with a physics engine and analyzed its results with SIR graphs. We found clear evidence of the relation between social distancing and getting infected rates. We further provide evidence that multi agent cooperation may scale better with increasing environment complexity and lead to a behavior that closer to far more human behavior. Ya da maskenin takılma yüzdesi ne kadar etkiliyor bunu yazıcaz.

<https://www.scribbr.co.uk/thesis-dissertation/abstract/>

# INTRODUCTION

Ever since the outbreak of Severe Acute Respiratory Syndrome as known as Covid-19 came out, life has changed drastically. Many researchers dedicated themselves to fight against the spread of this fatal virus and minimize the loss. Artificial Intelligence researchers are focusing their expertise knowledge to develop mathematical models for analyzing this epidemic disease.[1] An epidemic disease requires quick decisions to be made about interventions that could reduce or contain the disease spread. Decision makers need to be agile about their strategy since they race with the time. Every second that is wasted doubles the damage to humanity. In contrast, in order to decide confidently which strategy will work, decision makers need to analyze many scenarios and variables since if they give a wrong decision, that can also cause harm. This is an optimization problem which researchers cope by creating their own data and utilizing Reinforcement Learning to develop optimal strategies.[2]

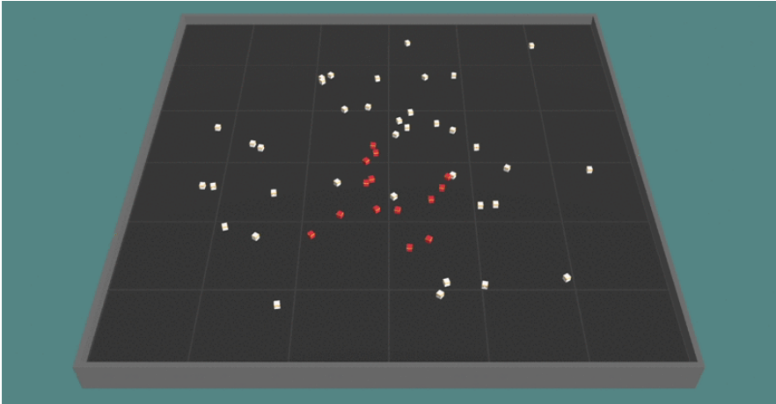
Reinforcement Learning is an area of Machine Learning where an agent learns the best behavior by interacting with the environment. Creating these complex environments and artificial intelligent agents that solve complex human relevant tasks has been life-long challenge for RL researchers.[3] Environment which is a crucial component of RL, describes the task that the agent attempts to solve. Agent and environments cannot be considered separately, and it is only a design choice for the researcher to determine where the environment starts and the agent ends. Most of the time environment is defined as anything that agent cannot have control on directly. In this article we introduce an environment which can simulate the epidemic spread with physics engine of Unity. With the help of Reinforcement Learning, agents are trained to learn social distancing by their own.

Figure 1:The Environment and agents. The square shape court limits the area and the cubes are agents. White ones represent healthy and red ones represent infectious

Precautions which became part of daily life can be given as an example of complex human-relevant tasks for Reinforcement Learning. Being one of the most common precautions, social distancing is a successful strategy to prevent the infectious disease from spreading. It has many forms but at its core, the aim is to keep people apart enough from each other by putting physical distance and/or confining them to their homes.

Creating intelligent artificial agents that can solve a wide variety of complex human-relevant tasks has been a long-standing challenge in the artificial intelligence community. Of particular relevance to humans will be agents that can sense and interact with objects in a physical world. One approach to creating these agents is to explicitly specify desired tasks and train a reinforcement learning (RL) agent to solve them. On this front, there has been much recent progress in solving physically grounded tasks, e.g. dexterous in-hand manipulation [(Rajeswaran et al.](#_bookmark58), [2017](#_bookmark58); [Andrychowicz et al.](#_bookmark14), [2018](#_bookmark14)) or locomotion of complex bodies ([Schulman et al.](#_bookmark62), [2015](#_bookmark62); [Heess et al.](#_bookmark34), [2017](#_bookmark34)). However, specifying reward functions or collecting demonstrations in order to supervise these tasks can be

for adaptation. These evolutionary arms races create implicit *autocurricula* ([Leibo et al.](#_bookmark44), [2019a](#_bookmark44)) whereby competing agents continually create new tasks for each other. There has been much success in leveraging multi-agent autocurricula to solve multi-player games, both in classic discrete games such as Backgammon ([Tesauro](#_bookmark79), [1995](#_bookmark79)) and Go ([Silver et al.](#_bookmark66), [2017](#_bookmark66)), as well as in continuous real-time domains such as Dota ([OpenAI](#_bookmark51), [2018](#_bookmark51)) and Starcraft ([Vinyals et al.](#_bookmark82), [2019](#_bookmark82)). Despite the impressive emergent complexity in these environments, the learned behavior is quite abstract and disembodied from the physical world. Our work sees itself in the tradition of previous studies that showcase emergent complexity in simple physically grounded environments ([Sims](#_bookmark67), [1994a](#_bookmark67); [Bansal et al.](#_bookmark17), [2018](#_bookmark17); [Jaderberg et al.](#_bookmark40), [2019](#_bookmark40); [Liu et al.](#_bookmark46), [2019](#_bookmark46)); the success in these settings inspires confidence that inducing autocurricula in physically grounded and open-ended environments could eventually enable agents to acquire an unbounded number of human-relevant skills.

We introduce a new mixed competitive and cooperative physics-based environment in which agents compete in a simple game of hide-and-seek. Through only a visibility-based reward function and competition, agents learn many emergent skills and strategies including collaborative tool use, where agents intentionally change their environment to suit their needs. For example, hiders learn to create shelter from the seekers by barricading doors or constructing multi-object forts, and as a counter strategy seekers learn to use ramps to jump into hiders’ shelter. Moreover, we observe signs of dy- namic and growing complexity resulting from multi-agent competition and standard reinforcement learning algorithms; we find that agents go through as many as six distinct adaptations of strategy and counter-strategy, which are depicted in Figure [1](#_bookmark0). We further present evidence that multi-agent co-adaptation may scale better with environment complexity and qualitatively centers around more human-interpretable behavior than intrinsically motivated agents.

However, as environments increase in scale and multi-agent autocurricula become more open-ended, evaluating progress by qualitative observation will become intractable. We therefore propose a suite of targeted intelligence tests to measure capabilities in our environment that we believe our agents may eventually learn, e.g. object permanence ([Baillargeon & Carey](#_bookmark16), [2012](#_bookmark16)), navigation, and construction. We find that for a number of the tests, agents pretrained in hide-and-seek learn faster or achieve higher final performance than agents trained from scratch or pretrained with intrinsic motivation; however, we find that the performance differences are not drastic, indicating that much of the skill and feature representations learned in hide-and-seek are entangled and hard to fine-tune.

The main contributions of this work are: 1) clear evidence that multi-agent self-play can lead to emergent autocurricula with many distinct and compounding phase shifts in agent strategy, 2) evi- dence that when induced in a physically grounded environment, multi-agent autocurricula can lead to human-relevant skills such as tool use, 3) a proposal to use transfer as a framework for evaluating agents in open-ended environments as well as a suite of targeted intelligence tests for our domain, and 4) open-sourced environments and code[1](#_bookmark1) for environment construction to encourage further research in physically grounded multi-agent autocurricula.

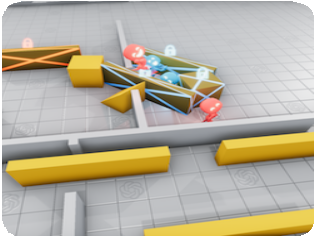
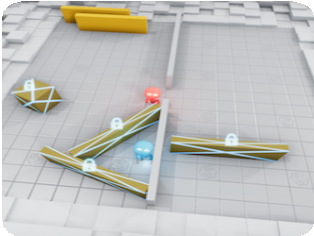
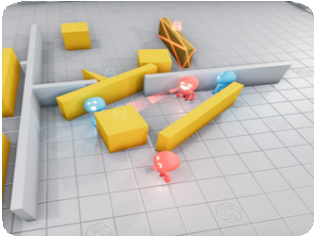
* 1. Running and Chasing (b) Fort Building (c) Ramp

Figure 1: Emergent Skill Progression From Multi-Agent Autocurricula. Through the reward signal of hide-and-seek (shown on the y-axis), agents go through 6 distinct stages of emergence. (a) Seekers (red) learn to chase hiders, and hiders learn to crudely run away. (b) Hiders (blue) learn basic tool use, using boxes and sometimes existing walls to construct forts. (c) Seekers learn to use ramps to jump into the hiders’ shelter. (d) Hiders quickly learn to move ramps to the edge of the play area, far from where they will build their fort, and lock them in place. (e) Seekers learn that they can jump from locked ramps to unlocked boxes and then *surf* the box to the hiders’ shelter, which is possible because the environment allows agents to move together with the box regardless of whether they are on the ground or not. (f) Hiders learn to lock all the unused boxes before constructing their fort. We plot the mean over 3 independent training runs with each individual seed shown with a dotted line. Please see [openai.com/blog/emergent-tool-use](https://openai.com/blog/emergent-tool-use) for example videos.

# RELATED WORK

Early work explored

Was further explored

More recent work attempted

In the context of Multi agent

Multi agent (cooperation) -Food Collection

Pandemic Simulation -SIR Model RL in Covid

Flocking- gibi başka distance ayarlama algoritmaları

# Flocking

# METHODOLOGY

# SOCIAL DISTANCING

# THEORETICAL FRAMEWORK

# RESEARCH DESIGN

# POLICY OPTIMIZATION

# RESULTS

# DISCUSSION AND CONCLUSION

# REFERENCES

[1] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, “COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms,” *medRxiv*, p. 2020.04.08.20057679, 2020.

[2] A. Yañez, C. Hayes, and F. Glavin, “Towards the control of epidemic spread: Designing reinforcement learning environments,” *CEUR Workshop Proc.*, vol. 2563, pp. 188–199, 2019.

[3] B. Baker *et al.*, “Emergent Tool Use From Multi-Agent Autocurricula,” 2019.

# Appendix