SSM Course 2023-24

MARL Systems

Mirco Musolesi

mircomusolesi@acm.org

Definition of Multiagent Systems

- Several possible definitions:
  - *Multiagent systems are distributed systems of independent actors called agents that are each independently controlled but that interact with one another in the same environment.* (see: Wooldridge, “Introduction to Multiagent Systems”, 2002 and Tulys and Stone, “Multiagent Learning Paradigms”, 2018).
  - *Multiagent systems are systems that include multiple autonomous entities with (possibly) diverging information* (see Shoham and Leyton-Brown, “Multiagent systems”, 2009).

Definition of Multiagent Learning

- We will use the following definition of multiagent learning:
  - *"The study of multiagent systems in which one or more of the autonomous entities improves automatically through experience*".

Characteristics of Multiagent Learning

- Different scale:
  - A city or an ant colony or a football team.
- Different degree of complexity:
  - A human, a machine, a mammal or an insect.
- Different types of interaction:
  - Frequent interactions (or not), interactions with a limited number of individuals, etc.

Prisoners' Dilemma

	Defect	Cooperate
Defect	(1, 1)	(10, 0)
Cooperate	(0, 10)	(5, 5)

Example: Prisoner's Dilemma

- Normal games were initially introduced as one-shot game.
- The players know each other's full utility (reward) functions and play the game only once.
- In this setting, the concept of Nash equilibrium was introduced: a set of actions such that no player would be better off deviating given that the other player's actions are fixed.
- Games can have one or multiple Nash equilibria.
- In the Prisoner's Dilemma, the only Nash Equilibrium is for both agents to defect.

[13] Whitehead, J. H. C., "Simple Homotopy Types." If $W = 1$, Theorem 5 follows from (17:3) on p. 155 of S. Lefschetz, *Algebraic Topology*, (New York, 1942) and arguments in §6 of J. H. C. Whitehead, "On Simply Connected 4-Dimensional Polyhedra" (*Comm. Math. Helv.*, 22, 48–92 (1949)). However this proof cannot be generalized to the case $W \neq 1$.

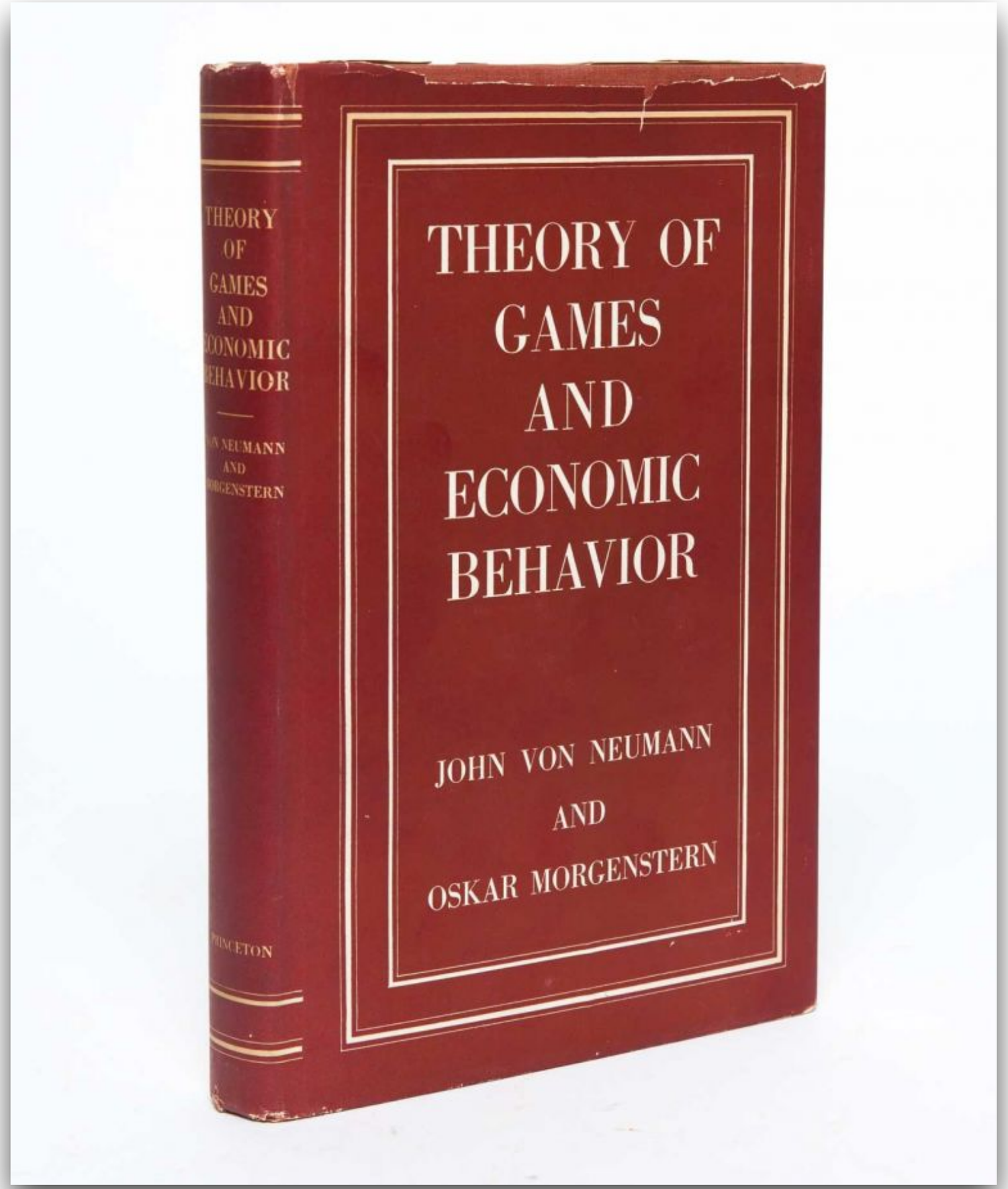
EQUILIBRIUM POINTS IN N-PERSON GAMES

By John F. Nash, Jr.*

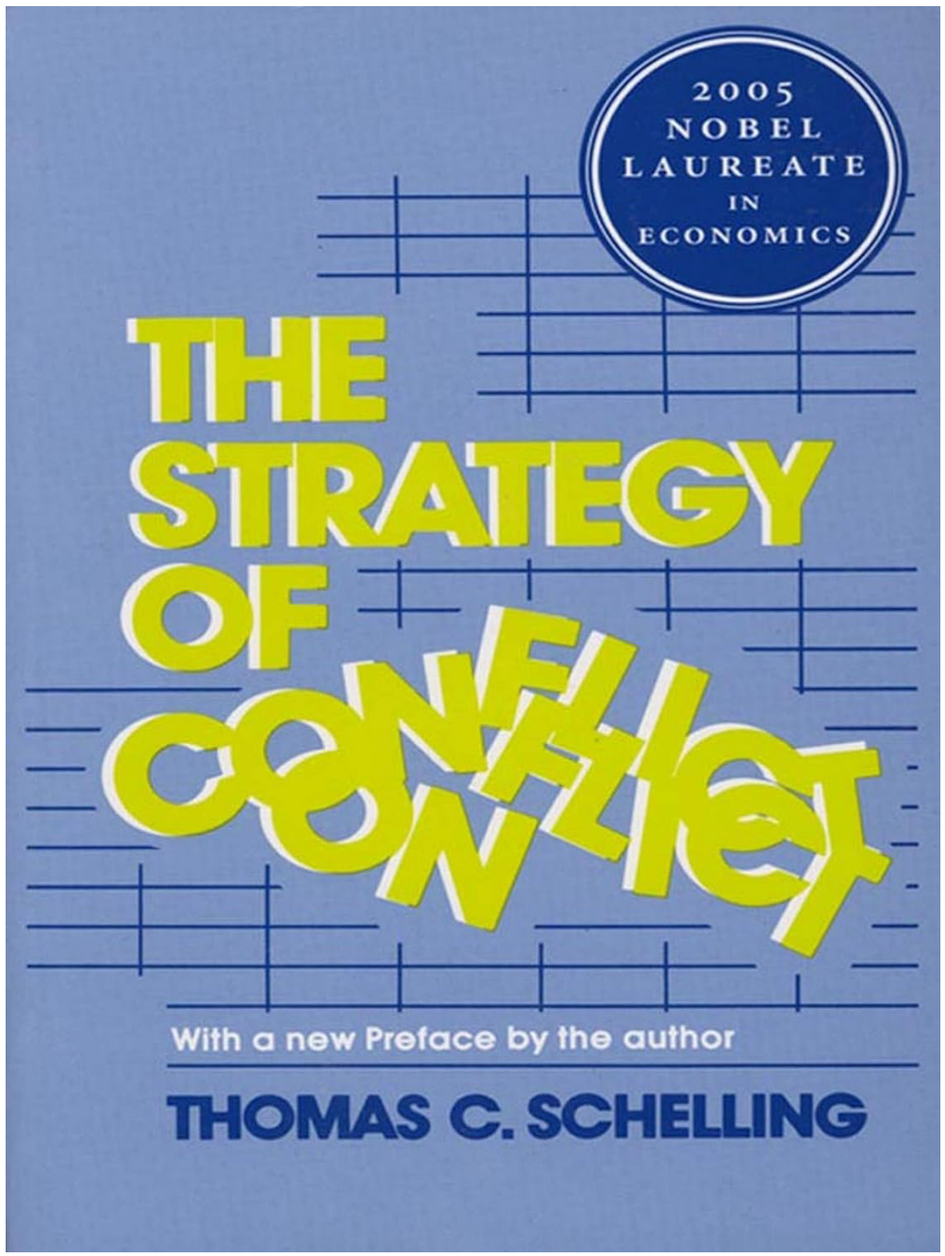
Princeton University

Communicated by S. Lefschetz, November 16, 1949

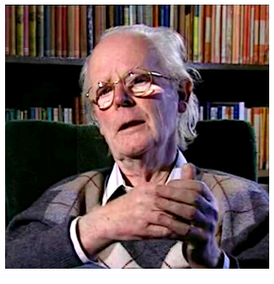
One may define a concept of an n-person game in which each player has a finite set of pure strategies and in which a definite set of payments to the n players corresponds to each n-tuple of pure strategies, one strategy being taken for each player. For mixed strategies, which are probability


THEORY OF GAMES AND ECONOMIC BEHAVIOR
JOHN VON NEUMANN AND OSKAR MORGENSTERN
THEORY OF GAMES AND ECONOMIC BEHAVIOR
N NEUMANN AND RGENSTERN
PRINCETON

Strategic Decision-making



Modelling Decision-making and Games



The Logic of Animal Conflict

J. MAYNARD SMITH
School of Biological Sciences, University of Sussex, Falmer, Sussex BN1 9QG

G. R. PRICE
Galton Laboratory, University College London, 4 Stephenson Way, London NW1 2HE

Conflicts between animals of the same species usually are of "limited war" type, not causing serious injury. This is often explained as due to group or species selection for behaviour benefiting the species rather than individuals. Game theory and computer simulation analyses show, however, that a "limited war" strategy benefits individual animals as well as the species.

IN a typical combat between two male animals of the same species, the winner gains mates, dominance rights, desirable territory, or other advantages that will tend toward transmitting its genes to future generations at higher frequencies than the loser's genes. Consequently, one might expect that natural selection would develop maximally effective weapons and fighting styles for a "total war" strategy of battles between males to the death. But instead, intraspecific conflicts are usually of a "limited war" type, involving inefficient weapons or ritualized tactics that seldom cause serious injury to either contestant. For example, in many snake species the males fight each other by wrestling without using their fangs[1,2]. In mule deer (*Odocoileus*

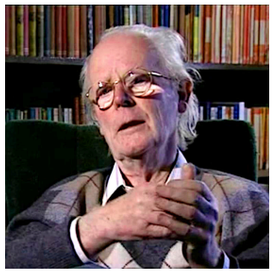
and ask what strategy will be favoured under individual selection. We first consider conflict in species possessing offensive weapons capable of inflicting serious injury on other members of the species. Then we consider conflict in species where serious injury is impossible, so that victory goes to the contestant who fights longest. For each model, we seek a strategy that will be stable under natural selection; that is, we seek an "evolutionarily stable strategy" or ESS. The concept of an ESS is fundamental to our argument; it has been derived in part from the theory of games, and in part from the work of MacArthur[13] and of Hamilton[14] on the evolution of the sex ratio. Roughly, an ESS is a strategy such that, if most of the members of a population adopt it, there is no "mutant" strategy that would give higher reproductive fitness.

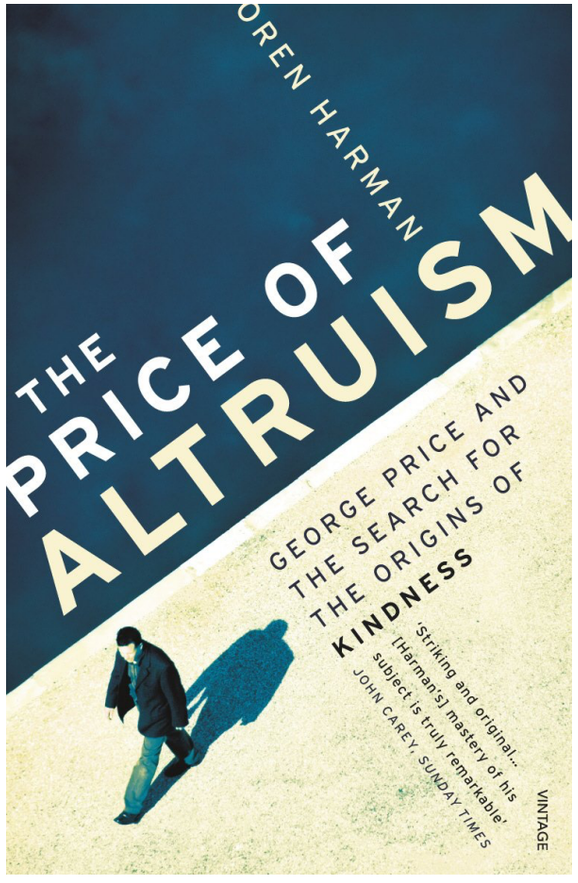
A Computer Model

A main reason for using computer simulation was to test whether it is possible even in theory for individual selection to account for "limited war" behaviour.

We consider a species that possesses offensive weapons capable of inflicting serious injuries. We assume that there are two categories of conflict tactics: "conventional" tactics, *C*, which are unlikely to cause serious injury, and "dangerous" tactics, *D*, which are likely to injure the opponent seriously if they are employed for long. (Thus in the snake example, wrestling involves *C* tactics and use of fangs would be *D* tactics. In many species *C* tactics

Modelling Decision-making and Games





The Logic of Animal Conflict

J. MAYNARD SMITH
School of Biological Sciences, University of Sussex, Falmer, Sussex BN1 9QG
G. R. PRICE
Galton Laboratory, University College London, 4 Stephenson Way, London NW1 2HE

Conflicts between animals of the same species usually are of "limited war" type, not causing serious injury. This is often explained as due to group or species selection for behaviour benefiting the species rather than individuals. Game theory and computer simulation analyses show, however, that a "limited war" strategy benefits individual animals as well as the species.

IN a typical combat between two male animals of the same species, the winner gains mates, dominance rights, desirable territory, or other advantages that will tend toward transmitting its genes to future generations at higher frequencies than the loser's genes. Consequently, one might expect that natural selection would develop maximally effective weapons and fighting styles for a "total war" strategy of battles between males to the death. But instead, intraspecific conflicts are usually of a "limited war" type, involving inefficient weapons or ritualized tactics that seldom cause serious injury to either contestant. For example, in many snake species the males fight each other by wrestling without using their fangs[1,2]. In mule deer (*Odocoileus*

and ask what strategy will be favoured under individual selection. We first consider conflict in species possessing offensive weapons capable of inflicting serious injury on other members of the species. Then we consider conflict in species where serious injury is impossible, so that victory goes to the contestant who fights longest. For each model, we seek a strategy that will be stable under natural selection; that is, we seek an "evolutionarily stable strategy" or ESS. The concept of an ESS is fundamental to our argument; it has been derived in part from the theory of games, and in part from the work of MacArthur[13] and of Hamilton[14] on the evolution of the sex ratio. Roughly, an ESS is a strategy such that, if most of the members of a population adopt it, there is no "mutant" strategy that would give higher reproductive fitness.

A Computer Model

A main reason for using computer simulation was to test whether it is possible even in theory for individual selection to account for "limited war" behaviour.

We consider a species that possesses offensive weapons capable of inflicting serious injuries. We assume that there are two categories of conflict tactics: "conventional" tactics, *C*, which are unlikely to cause serious injury, and "dangerous" tactics, *D*, which are likely to injure the opponent seriously if they are employed for long. (Thus in the snake example, wrestling involves *C* tactics and use of fangs would be *D* tactics. In many species *C* tactics

Strategic Decision-making

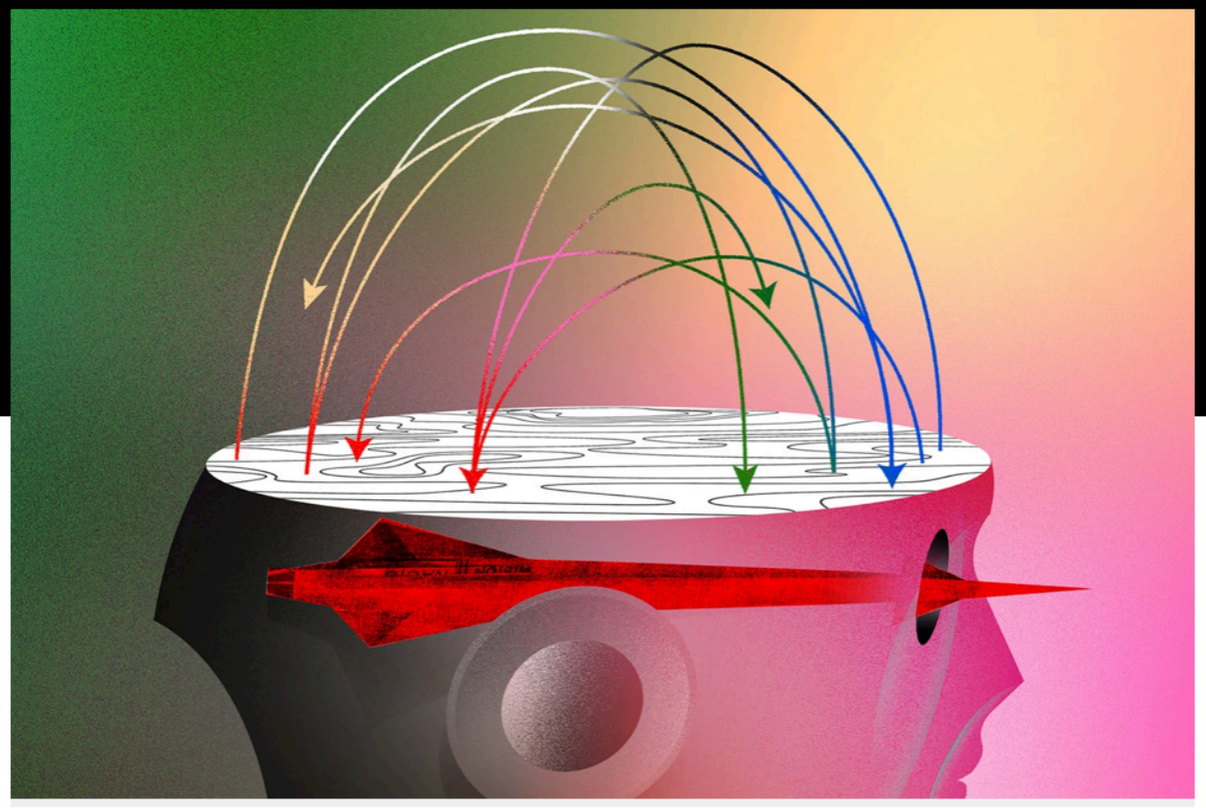
ANALYSIS

AI Has Entered the Situation Room

Data lets us see with unprecedented clarity—but reaping its benefits requires changing how foreign policy is made.

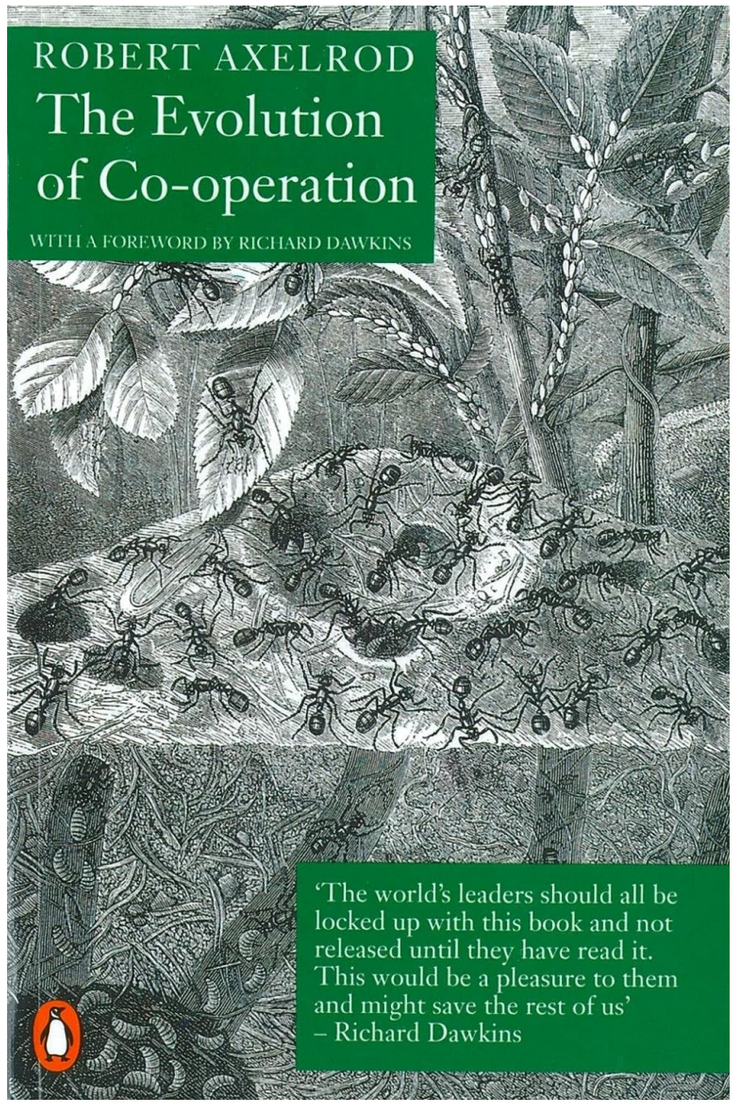
JUNE 19, 2023, 11:00 PM

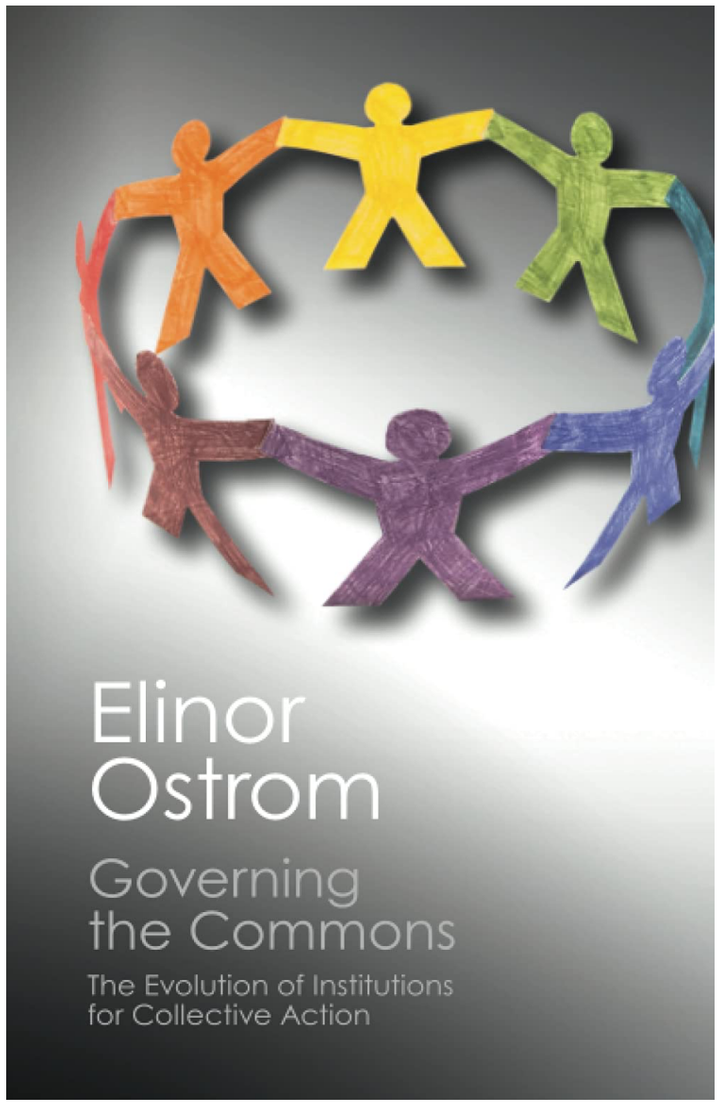
By Stanley McChrystal, a retired four-star U.S. Army general and an advisor to Rhombus Power, and Anshu Roy, the founder and CEO of Rhombus Power.



BRIAN STAUFFER ILLUSTRATION FOR FOREIGN POLICY

Challenge: Cooperative AI





[Nicolas Anastassacos, Stephen Hailes and Mirco Musolesi. Partner Selection for the Emergence of Cooperation in Multi-Agent Systems Using Reinforcement Learning. In *Proceedings of AAAI'20*. February 2020.]

Challenge:
the Tragedy of the AI Commons

The Tragedy of the Commons

The population problem has no technical solution;
it requires a fundamental extension in morality.

Garrett Hardin

At the end of a thoughtful article on the future of nuclear war, Wiesner and York (*1*) concluded that: "Both sides in the arms race are . . . confronted by the dilemma of steadily increasing military power and steadily decreasing national security. *It is our considered professional judgment that this dilemma has no technical solution.* If the great powers continue to look for solutions in the area of science and technology only, the result will be to worsen the situation."

I would like to focus your attention not on the subject of the article (national security in a nuclear world) but on the kind of conclusion they reached, namely that there is no technical solution to the problem. An implicit and almost universal assumption of discussional judgment. . . ." Whether they were right or not is not the concern of the present article. Rather, the concern here is with the important concept of a class of human problems which can be called "no technical solution problems," and, more specifically, with the identification and discussion of one of these.

It is easy to show that the class is not a null class. Recall the game of tick-tack-toe. Consider the problem, "How can I win the game of tick-tack-toe?" It is well known that I cannot, if I assume (in keeping with the conventions of game theory) that my opponent understands the game perfectly. Put another way, there is no "technical solution" to the problem. I can win only by giving a radical meaning to the word "win." I can hit my opponent over the

What Shall We Maximize?

Population, as Malthus said, naturally tends to grow "geometrically," or, as we would now say, exponentially. In a finite world this means that the per capita share of the world's goods must steadily decrease. Is ours a finite world?

A fair defense can be put forward for the view that the world is infinite; or that we do not know that it is not. But, in terms of the practical problems that we must face in the next few generations with the foreseeable technology, it is clear that we will greatly increase human misery if we do not, during the immediate future, assume that the world available to the terrestrial human population is finite. "Space" is no escape (*2*).

A finite world can support only a finite population; therefore, population growth must eventually equal zero. (The case of perpetual wide fluctuations above and below zero is a trivial variant that need not be discussed.) When this condition is met, what will be the situation of mankind? Specifically, can Bentham's goal of "the greatest good for the greatest number" be realized?

No—for two reasons, each sufficient by itself. The first is a theoretical one. It is not mathematically possible to maximize for two (or more) variables at the same time. This was clearly stated by von Neumann and Morgenstern (*3*), but the principle is implicit in the theory of partial differential equations, dating

DeepMind

2020-12-17

Open Problems in Cooperative AI

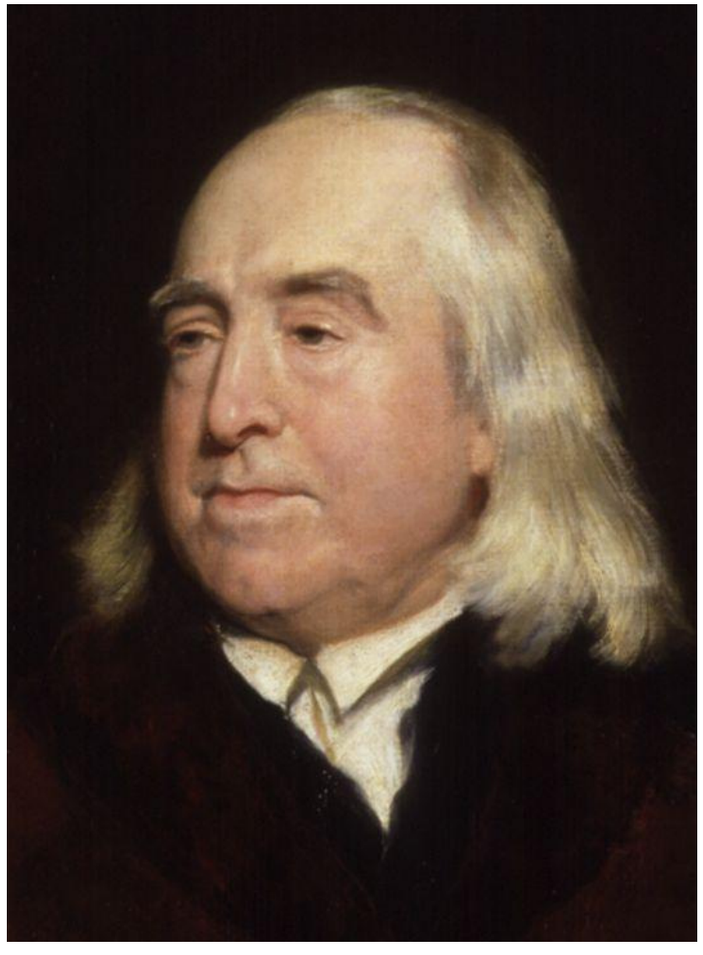
Allan Dafoe[1], Edward Hughes[2], Yoram Bachrach[2], Tantum Collins[2], Kevin R. McKee[2], Joel Z. Leibo[2], Kate Larson[2, 3] and Thore Graepel[2]

[1]Centre for the Governance of AI, Future of Humanity Institute, University of Oxford, [2]DeepMind, [3]University of Waterloo

Problems of cooperation—in which agents seek ways to jointly improve their welfare—are ubiquitous and important. They can be found at scales ranging from our daily routines—such as driving on highways, scheduling meetings, and working collaboratively—to our global challenges—such as peace, commerce, and pandemic preparedness. Arguably, the success of the human species is rooted in our ability to cooperate. Since machines powered by artificial intelligence are playing an ever greater role in our lives, it will be important to equip them with the capabilities necessary to cooperate and to foster cooperation.

Challenge:
Ethics and Decision-making

Credit: Wikimedia



Credit: Wikimedia

Credit: Wikimedia

[Elizaveta Tennant, Stephen Hailes and Mirco Musolesi. Modeling Moral Choices in Social Dilemmas with Multi-agent Reinforcement Learning. In Proceedings of the 32nd Joint Conference on Artificial Intelligence (IJCAI 2023). August 2023.]

RLQ: Workload Allocation with Reinforcement Learning in Distributed Queues

Alessandro Staffolani, Victor-Alexandru Darvariu, Paolo Bellavista and Mirco Musolesi

Abstract—Distributed workload queues are nowadays widely used due to their significant advantages in terms of decoupling, resilience, and scaling. Task allocation to worker nodes in distributed queue systems is typically simplistic (e.g., Least Recently Used) or uses hand-crafted heuristics that require task-specific information (e.g., task resource demands or expected time of execution). When such task information is not available and worker node capabilities are not homogeneous, the existing placement strategies may lead to unnecessarily large execution timings and usage costs. In this work, we investigate the task allocation problem within the *Markov Decision Process* framework, where an agent assigns tasks to an available resource, by receiving a numerical reward signal upon task completion. This allows our solution to learn effective task allocation strategies directly from experience in a completely dynamic way. In particular, we present the design, implementation, and experimental evaluation of *RLQ* (Reinforcement Learning based Queues), i.e., our adaptive and learning-based task allocation solution that we have implemented and integrated with the popular Celery task queuing system. By using both synthetic and real workload traces, we compare *RLQ* against traditional solutions, such as Least Recently Used. On average, using synthetic workloads, *RLQ* reduces the execution time by a factor of at least $3\times$. When considering the execution cost, the reduction is around 70%, whereas for the time waited before execution, the reduction is close to a factor of $7\times$. Using real traces, we observe around 70% improvement for execution time, around 20% for execution cost and a reduction of approximately $20\times$ for waiting time. We also analyze *RLQ* performance against E-PVM, a state-of-the-art solution used in Google's Borg, showing that we are able to outperform it in the synthetic data evaluation, while we outperform it in all the three settings based on real data.

Index Terms—task allocation, reinforcement learning, distributed task queuing.

1 Introduction

The problem of task scheduling concerns performing allocations to resources so as to satisfy desired, often conflicting, objectives (such as throughput, latency, or fairness) while accounting for underlying architectural properties. This problem presents itself at multiple levels in computer systems; notable examples include scheduling of threads on processors [1], [2], scheduling of packets in network infrastructure [3], [4], stream processing [5], [6], software cache and information about the underlying hardware utilization is not available (or expensive to obtain). A relevant use case is that of federated cloud deployments, where the cloud infrastructure belonging to several owners is leased to a client in order to satisfy its business needs. In this situation, the infrastructure owners typically limit hardware monitoring. Another example is that of citizen science projects such as SETI@home [12] and Folding@home [13], in which

Control-Tutored Reinforcement Learning: Towards the Integration of Data-Driven and Model-Based Control

Francesco DeLellis FRANCESCO.DELELLIS@UNINA.IT
University of Naples Federico II, Italy

Marco Coraggio MARCO.CORAGGIO@UNINA.IT
Scuola Superiore Meridionale, Italy

Giovanni Russo* GIOVARUSSO@UNISA.IT
University of Salerno, Italy

Mirco Musolesi* M.MUSOLESI@UCL.AC.UK
University College London, UK, and University of Bologna, Italy

Mario di Bernardo* MARIO.DIBERNARDO@UNINA.IT
University of Naples Federico II, Italy, and Scuola Superiore Meridionale, Italy

Abstract

We present an architecture where a feedback controller derived on an approximate model of the environment assists the learning process to enhance its data efficiency. This architecture, which we term as Control-Tutored Q-Learning (CTQL), is presented in two alternative flavours. The former is based on defining the reward function so that a Boolean condition can be used to determine when the control tutor policy is adopted, while the latter, termed as probabilistic CTQL (pCTQL), is instead based on executing calls to the tutor with a certain probability during learning. Both approaches are validated, and thoroughly benchmarked against Q-Learning, by considering the stabilization of an inverted pendulum as defined in OpenAI Gym as a representative problem.

Keywords: Reinforcement learning based control, data-driven control, feedback control.

PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Darvariu V-A, Hailes S, Musolesi M. 2021 Goal-directed graph construction using reinforcement learning. *Proc. R. Soc. A* **477**: 20210168.
https://doi.org/10.1098/rspa.2021.0168

Received: 23 February 2021
Accepted: 29 September 2021

Subject Areas:
complexity, artificial intelligence

Keywords:
network robustness, complex systems, graph neural networks, reinforcement learning

Author for correspondence:

Goal-directed graph construction using reinforcement learning

Victor-Alexandru Darvariu[1,2], Stephen Hailes[1] and Mirco Musolesi[1,2,3]

[1]Department of Computer Science, University College London, London, UK
[2]The Alan Turing Institute, London, UK
[3]Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

V-AD, 0000-0001-9250-8175; MM, 0000-0001-9712-4090

Graphs can be used to represent and reason about systems and a variety of metrics have been devised to quantify their global characteristics. However, little is currently known about how to construct a graph or improve an existing one given a target objective. In this work, we formulate the construction of a graph as a decision-making process in which a central agent creates topologies by trial and error and receives rewards proportional to the value of the target objective. By means of this conceptual framework, we propose an algorithm based on reinforcement

Dynamic Network Reconfiguration for Entropy Maximization using Deep Reinforcement Learning

Christoffel Doorman[1], Victor-Alexandru Darvariu[1,2], Stephen Hailes[1], Mirco Musolesi[1,2,3]
[1]University College London [2]The Alan Turing Institute [3]University of Bologna
{christoffel.doorman.20, v.darvariu, s.hailes, m.musolesi}@ucl.ac.uk

Abstract

A key problem in network theory is how to reconfigure a graph in order to optimize a quantifiable objective. Given the ubiquity of networked systems, such work has broad practical applications in a variety of situations, ranging from drug and material design to telecommunications. The large decision space of possible reconfigurations, however, makes this problem computationally intensive. In this paper, we cast the problem of network rewiring for optimizing a specified structural property as a Markov Decision Process (MDP), in which a decision-maker is given a budget of modifications that are performed sequentially. We then propose a general approach based on the Deep Q-Network (DQN) algorithm and graph neural networks (GNNs) that can efficiently learn strategies for rewiring networks. We then discuss a cybersecurity case study, i.e., an application to the computer network reconfiguration problem for intrusion protection. In a typical scenario, an attacker might have a (partial) map of the system they plan to penetrate; if the network is effectively "scrambled", they would not be able to navigate it since their prior knowledge would become obsolete. This can be viewed as an entropy maximization problem, in which the goal is to increase the *surprise* of the network. Indeed, entropy acts as a proxy measurement of the difficulty of navigating the network topology. We demonstrate the general ability of the proposed method to obtain better entropy gains than random rewiring on synthetic and real-world graphs while being computationally inexpensive, as well as being able to generalize to larger graphs than those seen during training. Simulations of attack scenarios confirm the effectiveness of the learned rewiring strategies.

References

- Stefano V. Albrecht and Peter Stone. Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems. Artificial Intelligence. Volume 258. 2018.

- Lucian Busoniu, Robert Babuska, Bart De Schutter. A Comprehensive Survey of Multiagent Reinforcement Learning. In IEEE Transactions on Systems, Man and Cybernetics. Volume 38. Issue 2. March 2008.

References

- Kevin Leyton-Brown and Yoav Shoham. Multiagent Systems, Game-theoretic and Logical Foundations. Cambridge University Press. 2009.

- Karl Tuyls and Gerhard Weiss. Multiagent Learning: Basics, Challenges and Prospects. AI Magazine. Volume 33. Issue 3. 2012.

References

- Karl Tulys and Peter Stone. Multiagent Learning Paradigms. In Francesco Belardinelli and Estefania Argente, editors, Multi-agent Systems and Agreement Technologies. Lecture Notes in Artificial Intelligence. Pages 3-21. Springer 2018.

- Micheal Wooldridge. An Introduction to MultiAgent Systems. Second Edition. Wiley. 2009.

References

- Kaiqing Zhang, Zhuoran Yang and Tamer Basar. Multi-agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. arXiv:1911.10635v2. 2021.

- Sven Gronauer and Klaus Diepold. Multi-agent Reinforcement Learning: A Survey. Artificial Intelligence Review. 55:895-943. Springer. 2022.