Towards a unified treatment of the dynamics of collective learning

Wolfram Barfuss

Challenges and Opportunities for Multi-Agent Reinforcement Learning - COMARL AAAI 2021

March 22-23, 2021

Max-Planck-Institut für Mathematik in den Naturwissenschaften

UNIVERSITY OF LEEDS

TÜBINGEN AI CENTER

Collective Learning Dynamics

An unusual presentation - no results, only questions

Collective Learning Dynamics

An unusual presentation - no results, only questions

- What do I mean with Collective Learning Dynamics?

Collective Learning Dynamics

An unusual presentation - no results, only questions

- What do I mean with Collective Learning Dynamics?
- Multiple literatures ➡ **can we learn from each other?**

Collective Learning Dynamics

An unusual presentation - no results, only questions

- What do I mean with Collective Learning Dynamics?
- Multiple literatures ➡ **can we learn from each other?**
- **How to use our** (unified) **knowledge**
  to help solving challenges of multi-agent reinforcement learning?

Collective Learning Dynamics

An unusual presentation - no results, only questions

- What do I mean with Collective Learning Dynamics?
- Multiple literatures ➡ **can we learn from each other?**
- **How to use our** (unified) **knowledge**
  to help solving challenges of multi-agent reinforcement learning?

My wish

Follow-up interactions & collaborations

Learning dynamics?

Learning dynamics?

=> Replicator Reinforcement Learning Dynamics

based on the link between evolutionary game theory and reinforcement learning

Learning dynamics?

Stochastic RL algorithms

⬇

=> **Replicator Reinforcement Learning Dynamics** = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Learning dynamics?

Stochastic RL algorithms

⬇

=> **Replicator Reinforcement Learning Dynamics** = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example

Emergence of cooperation

Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

A tragedy of the commons

Agent 1

Agent 2

	Co-operate	**D**efect
C	1, 1	-2, 3
D	3, -2	0, 0

Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

A tragedy of the commons

Agent 1

Agent 2

	Co-operate	**D**efect
C	1, 1	-2, 3
D	3, -2	0, 0

Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

A tragedy of the commons

Agent 1

Agent 2

	Co-operate	**D**efect
C	1, 1	-2, 3
D	3, -2	0, 0

Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

A tragedy of the commons

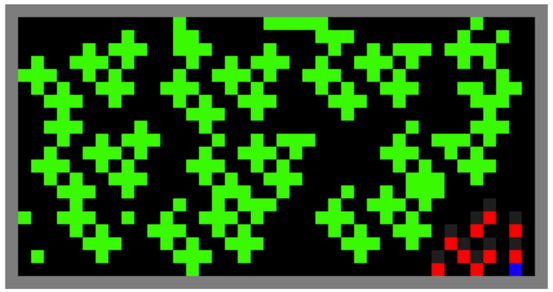
Agent 1

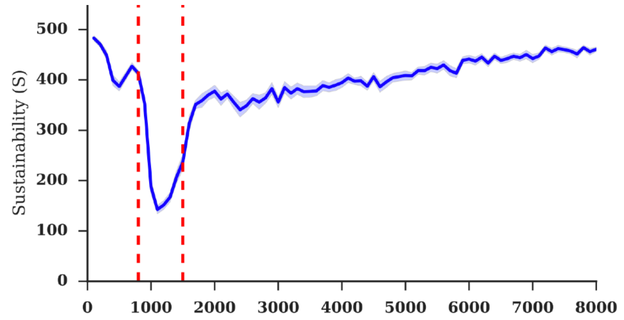
Agent 2

	Co-operate	**D**efect
C	1, 1	-2, 3
D	3, -2	0, 0

Deep MARL for common-pool resource appropriation

Perolat et al, 2017





Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

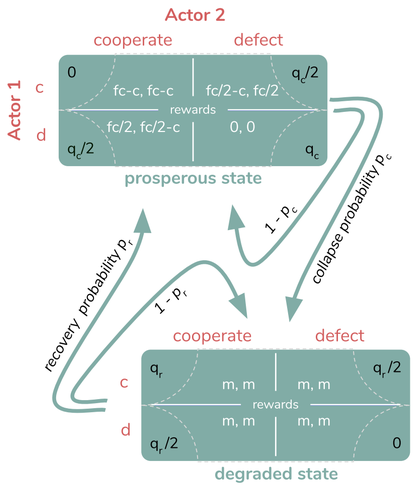
A tragedy of the commons

Agent 1

Agent 2	Co-operate	Defect
C	1, 1	-2, 3
D	3, -2	0, 0

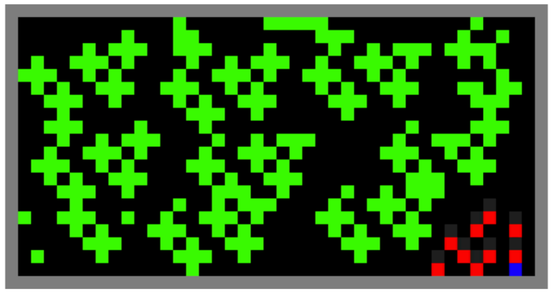
Social-ecological dilemma

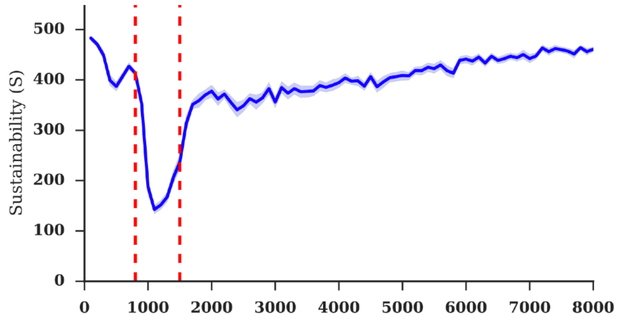
Barfuss et al., 2020



Deep MARL for common-pool resource appropriation

Perolat et al, 2017





Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

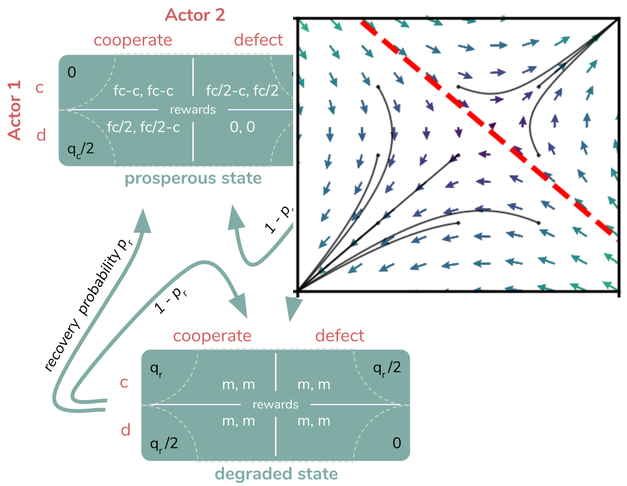
A tragedy of the commons

Agent 1

Agent 2	Co-operate	Defect
C	1, 1	-2, 3
D	3, -2	0, 0

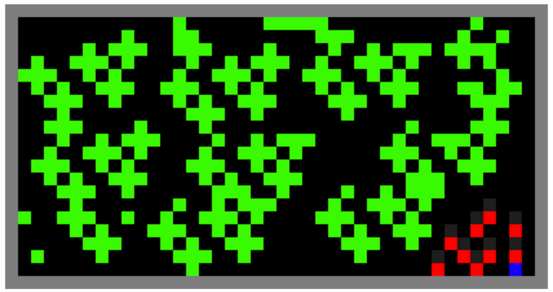
Social-ecological dilemma

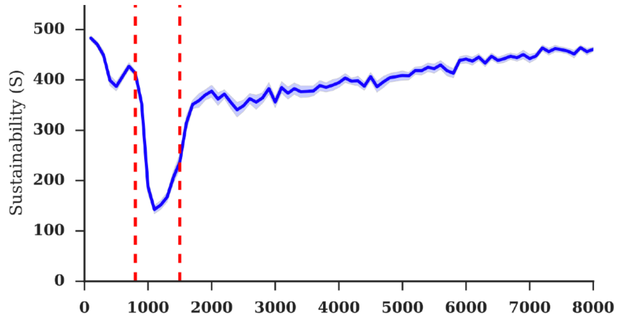
Barfuss et al., 2020



Deep MARL for common-pool resource appropriation

Perolat et al, 2017





Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

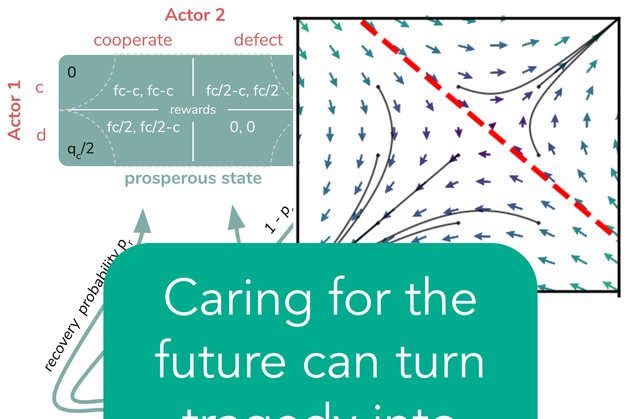
A tragedy of the commons

Agent 1

Agent 2	Co-operate	Defect
C	1, 1	-2, 3
D	3, -2	0, 0

Social-ecological dilemma

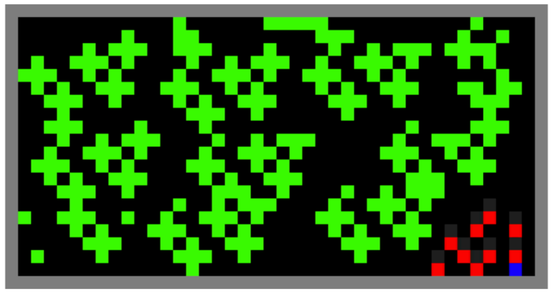
Barfuss et al., 2020

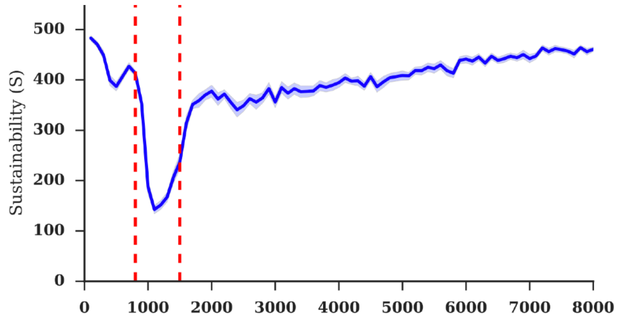


Caring for the future can turn tragedy into comedy

Deep MARL for common-pool resource appropriation

Perolat et al, 2017





Learning dynamics?

Stochastic RL algorithms

⬇

=> Replicator Reinforcement Learning Dynamics = Idealised learning equations

based on the link between evolutionary game theory and reinforcement learning

Example
Emergence of cooperation

Prisoner's Dilemma

A tragedy of the commons

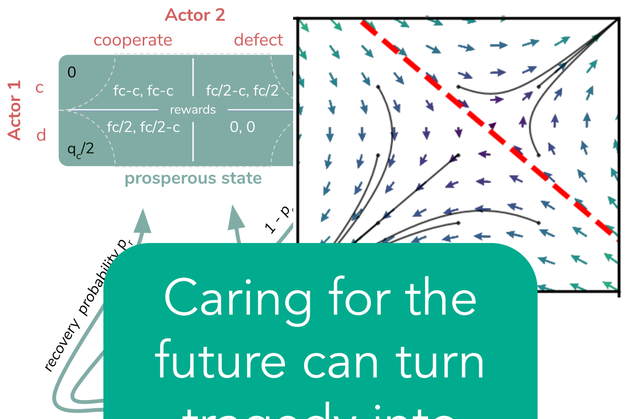
Agent 1

Agent 2

	Co-operate	**Defect**
C	1, 1	-2, 3
D	3, -2	0, 0

Social-ecological dilemma

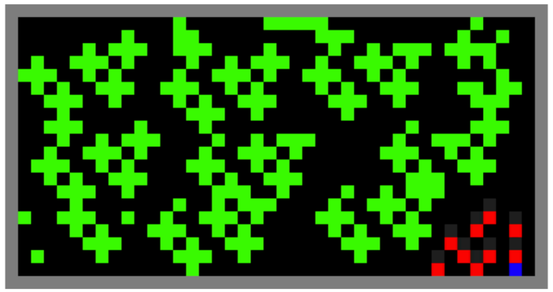
Barfuss et al., 2020

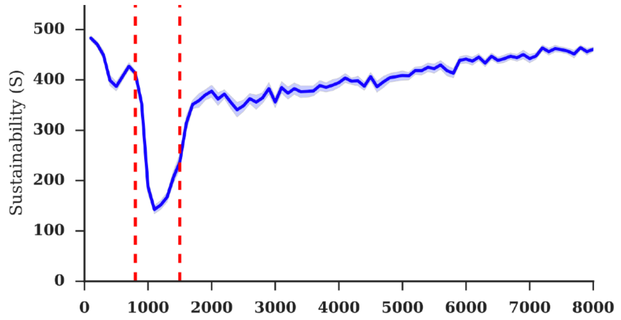


Caring for the future can turn tragedy into comedy

Deep MARL for common-pool resource appropriation

Perolat et al, 2017





➡ **a semi-formal method**

Why learning dynamics? ➡ Applications

Collective learning dynamics for solving MARL challenges

Why learning dynamics? ➡ Applications

Collective learning dynamics for solving MARL challenges

MARL Challenges

There has been a lot of great work on multi-agent reinforcement learning (MARL) in the past decade, but significant challenges remain, including:

- the difficulty of learning an optimal model/policy from a partial signal,
- learning to cooperate/compete in non-stationary environments with distributed, simultaneously learning agents,
- the interplay between abstraction and influence of other agents,
- the exploration vs. exploitation dilemma,
- the scalability and effectiveness of learning algorithms,
- avoiding social dilemmas, and
- learning emergent communication.

COMARL Symposium Website

Why learning dynamics? ➡ Applications

Collective learning dynamics for solving MARL challenges

MARL Challenges

There has been a lot of great work on multi-agent reinforcement learning (MARL) in the past decade, but significant challenges remain, including:

- the difficulty of learning an optimal model/policy from a partial signal,
- learning to cooperate/compete in non-stationary environments with distributed, simultaneously learning agents,
- the interplay between abstraction and influence of other agents,
- the exploration vs. exploitation dilemma,
- the scalability and effectiveness of learning algorithms,
- avoiding social dilemmas, and
- learning emergent communication.

COMARL Symposium Website

Application domains

- Interpretability
  of multi-agent learning systems
- Analysis of strategic interactions
  of multiple learning agents in changing environments
- Hyper-parameter tuning
- Design of new algorithms

Why learning dynamics? ➡ Applications

Collective learning dynamics for solving MARL challenges

MARL Challenges

There has been a lot of great work on multi-agent reinforcement learning (MARL) in the past decade, but significant challenges remain, including:

- the difficulty of learning an optimal model/policy from a partial signal,
- learning to cooperate/compete in non-stationary environments with distributed, simultaneously learning agents,
- the interplay between abstraction and influence of other agents,
- the exploration vs. exploitation dilemma,
- the scalability and effectiveness of learning algorithms,
- avoiding social dilemmas, and
- learning emergent communication.

COMARL Symposium Website

Application domains

- Interpretability
  of multi-agent learning systems
- Analysis of strategic interactions
  of multiple learning agents in changing environments
- Hyper-parameter tuning
- Design of new algorithms

Problem-Method Fit ⬅
How exactly?

The challenge

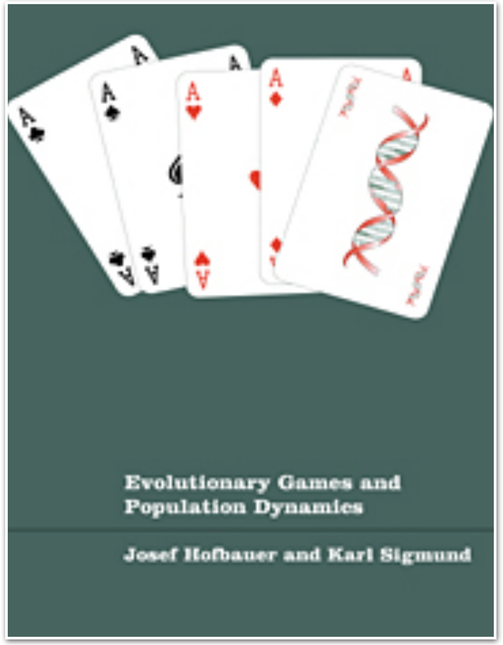
Disciplinary divides a clustering attempt

The challenge

Disciplinary divides a clustering attempt

Biology

Evolutionary Game Theory



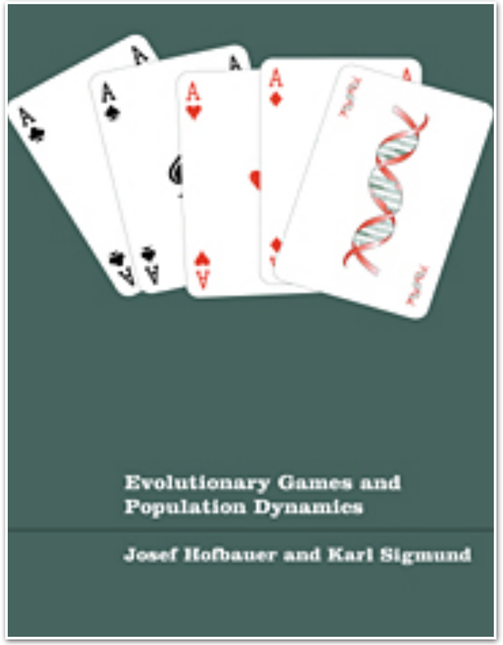
Hofbauer & Sigmund, 1998

The challenge

Disciplinary divides a clustering attempt

Biology

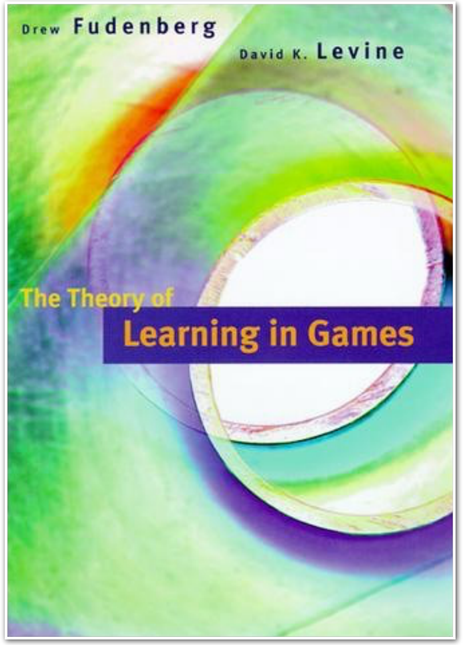
Evolutionary Game Theory



Hofbauer & Sigmund, 1998

Economics

Learning in Games



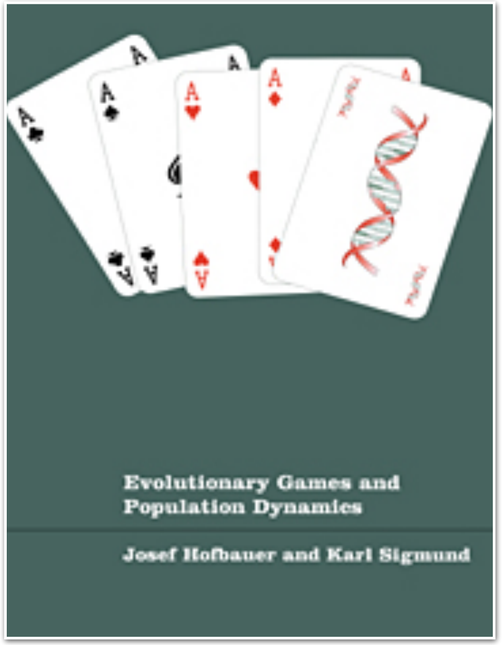
Fudenberg & Levine, 1998

The challenge

Disciplinary divides a clustering attempt

Biology

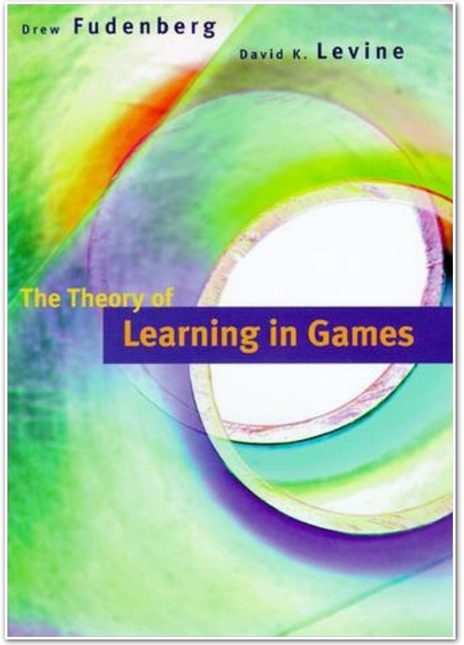
Evolutionary Game Theory



Hofbauer & Sigmund, 1998

Economics

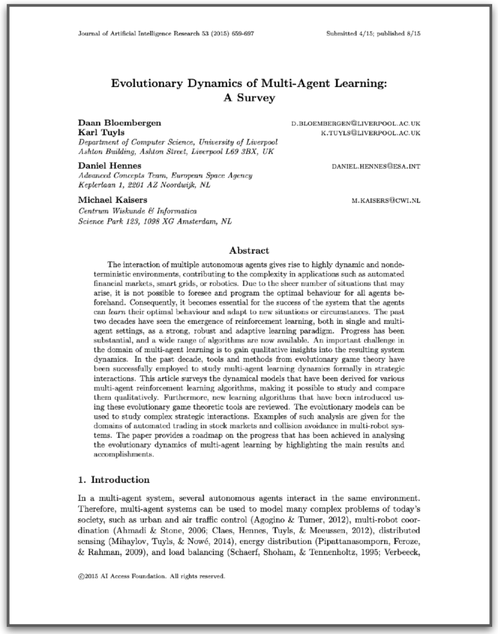
Learning in Games



Fudenberg & Levine, 1998

Computer Science

Machine Learning Dynamics



Journal of Artificial Intelligence Research 53 (2015) 659-697 Submitted 4/15; published 8/15

Evolutionary Dynamics of Multi-Agent Learning: A Survey

Daan Bloembergen D.BLOEMBERGEN@LIVERPOOL.AC.UK
Karl Tuyls K.TUYLS@LIVERPOOL.AC.UK
Department of Computer Science, University of Liverpool
Ashton Building, Ashton Street, Liverpool L69 3BX, UK

Daniel Hennes DANIEL.HENNES@ESA.INT
Advanced Concepts Team, European Space Agency
Keplerlaan 1, 2201 AZ Noordwijk, NL

Michael Kaisers M.KAISERS@CWI.NL
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, NL

Abstract

The interaction of multiple autonomous agents gives rise to highly dynamic and nondeterministic environments, contributing to the complexity in applications such as automated financial markets, smart grids, or robotics. Due to the sheer number of situations that may arise, it is not possible to foresee and program the optimal behaviour for all agents beforehand. Consequently, it becomes essential for the success of the system that the agents can *learn* their optimal behaviour and adapt to new situations or circumstances. The past two decades have seen the emergance of reinforcement learning, both in single and multi-agent settings, as a strong, robust and adaptive learning paradigm. Progress has been substantial, and a wide range of algorithms are now available. An important challenge in the domain of multi-agent learning is to gain qualitative insights into the resulting system dynamics. In the past decade, tools and methods from evolutionary game theory have been successfully employed to study multi-agent learning dynamics formally in strategic interactions. This article surveys the dynamical models that have been derived for various multi-agent reinforcement learning algorithms, making it possible to study and compare them qualitatively. Furthermore, new learning algorithms that have been introduced using these evolutionary game theoretic tools are reviewed. The evolutionary models can be used to study complex strategic interactions. Examples of such analysis are given for the domains of automated trading in stock markets and collision avoidance in multi-robot systems. The paper provides a roadmap on the progress that has been achieved in analysing the evolutionary dynamics of multi-agent learning by highlighting the main results and accomplishments.

1. Introduction

In a multi-agent system, several autonomous agents interact in the same environment. Therefore, multi-agent systems can be used to model many complex problems of today's society, such as urban and air traffic control (Agogino & Tumer, 2012), multi-robot coordination (Ahmadi & Stone, 2006; Claes, Hennes, Tuyls, & Meeussen, 2012), distributed sensing (Mihaylov, Tuyls, & Nowé, 2014), energy distribution (Pipattanasomporn, Feroze, & Rahman, 2009), and load balancing (Schaerf, Shoham, & Tennenholtz, 1995; Verbeeck,

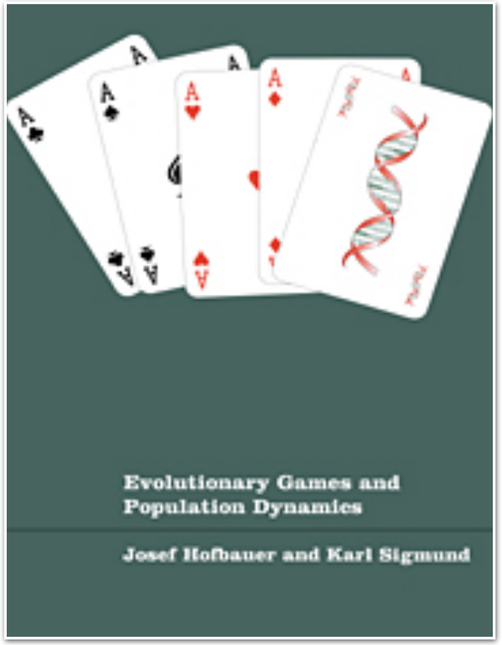
©2015 AI Access Foundation. All rights reserved.

Bloembergen et al., 2015

The challenge

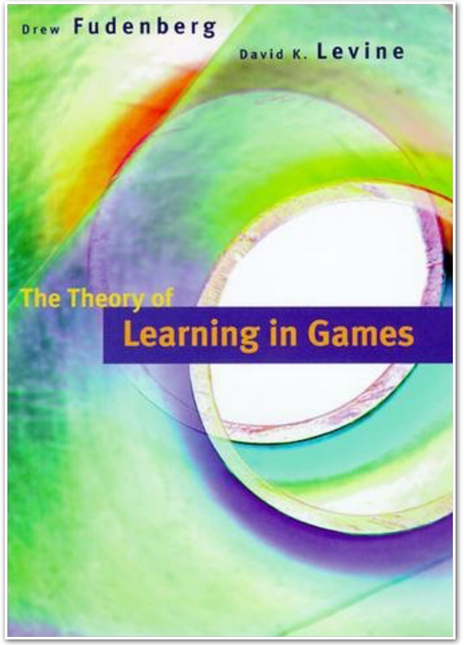
Disciplinary divides a clustering attempt

Biology
Evolutionary Game Theory



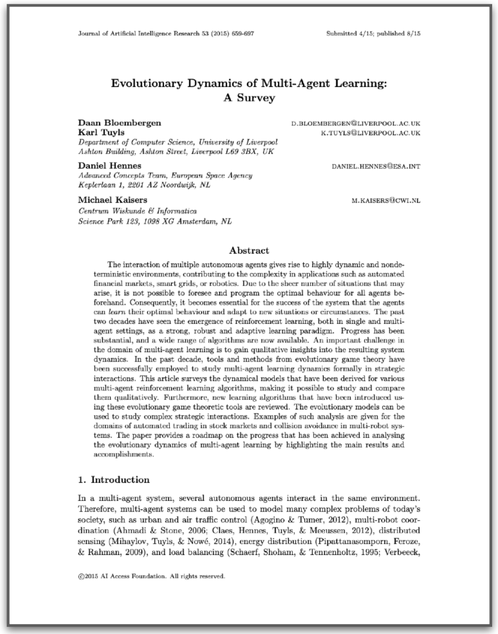
Hofbauer & Sigmund, 1998

Economics
Learning in Games



Fudenberg & Levine, 1998

Computer Science
Machine Learning Dynamics



Journal of Artificial Intelligence Research 53 (2015) 659-697 Submitted 4/15; published 8/15

Evolutionary Dynamics of Multi-Agent Learning: A Survey

Daan Bloembergen D.BLOEMBERGEN@LIVERPOOL.AC.UK
Karl Tuyls K.TUYLS@LIVERPOOL.AC.UK
Department of Computer Science, University of Liverpool
Ashton Building, Ashton Street, Liverpool L69 3BX, UK
Daniel Hennes DANIEL.HENNES@ESA.INT
Advanced Concepts Team, European Space Agency
Keplerlaan 1, 2201 AZ Noordwijk, NL
Michael Kaisers M.KAISERS@CWI.NL
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, NL

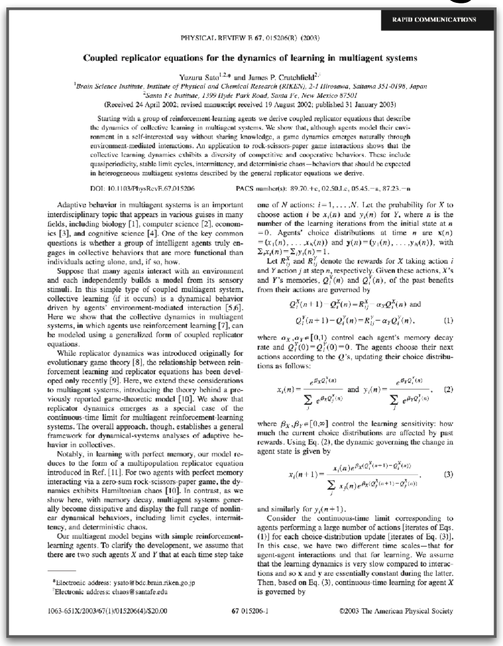
Abstract

The interaction of multiple autonomous agents gives rise to highly dynamic and nondeterministic environments, contributing to the complexity in applications such as automated financial markets, smart grids, or robotics. Due to the sheer number of situations that may arise, it is not possible to foresee and program the optimal behaviour for all agents beforehand. Consequently, it becomes essential for the success of the system that the agents can learn their optimal behaviour and adapt to new situations or circumstances. The past two decades have seen the emergence of reinforcement learning, both in single and multi-agent settings, as a strong, robust and adaptive learning paradigm. Progress has been substantial, and a wide range of algorithms are now available. An important challenge in the domain of multi-agent learning is to gain qualitative insights into the resulting system dynamics. In the past decade, tools and methods from evolutionary game theory have been successfully employed to study multi-agent learning dynamics formally in strategic interactions. This article surveys the dynamical models that have been derived for various multi-agent reinforcement learning algorithms, making it possible to study and compare them qualitatively. Furthermore, new learning algorithms that have been introduced using these evolutionary game theoretic tools are reviewed. The evolutionary models can be used to study complex strategic interactions. Examples of such analysis are given for the domains of automated trading in stock markets and collision avoidance in multi-robot systems. The paper provides a roadmap on the progress that has been achieved in analysing the evolutionary dynamics of multi-agent learning by highlighting the main results and accomplishments.

1. Introduction

In a multi-agent system, several autonomous agents interact in the same environment. Therefore, multi-agent systems can be used to model many complex problems of today's society, such as urban and air traffic control (Agogino & Turner, 2012), multi-robot coordination (Ahmadi & Stone, 2006; Claes, Hennes, Tuyls, & Meeussen, 2012), distributed sensing (Mihaylov, Tuyls, & Nowé, 2014), energy distribution (Pipattanasomporn, Feroze, & Rahman, 2009), and load balancing (Schaerf, Shoham, & Tennenholtz, 1995; Verbeeck,

Physics
Statistical mechanics learning



RAPID COMMUNICATIONS

PHYSICAL REVIEW E 67, 015206(R) (2003)

Coupled replicator equations for the dynamics of learning in multiagent systems

Yuzuru Sato[1,2,*] and James P. Crutchfield[2]
[1]*Brain Science Institute, Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Saitama 351-0198, Japan*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*
(Received 24 April 2002; revised manuscript received 19 August 2002; published 31 January 2003)

Starting with a group of reinforcement-learning agents we derive coupled replicator equations that describe the dynamics of collective learning in multiagent systems. We show that, although agents model their environment in a self-interested way without sharing knowledge, a game dynamics emerges naturally through environment-mediated interactions. An application to rock-scissors-paper game interactions shows that the collective learning dynamics exhibits a diversity of competitive and cooperative behaviors. These include quasiperiodicity, stable limit cycles, intermittency, and deterministic chaos—behaviors that should be expected in heterogeneous multiagent systems described by the general replicator equations we derive.

DOI: 10.1103/PhysRevE.67.015206 PACS number(s): 89.20.+a, 02.50.Le, 05.45.−a, 87.23.−n

Adaptive behavior in multiagent systems is an important interdisciplinary topic that appears in various guises in many fields, including biology [1], computer science [2], economics [3], and cognitive science [4]. One of the key common questions is whether a group of intelligent agents truly engages in collective behaviors that are more functional than individuals acting alone, and, if so, how.

Suppose that many agents interact with an environment and each independently builds a model from its sensory stimuli. In this simple type of coupled multiagent system, collective learning (if it occurs) is a dynamical behavior driven by agents' environment-mediated interaction [5,6]. Here we show that the collective dynamics in multiagent systems, in which agents use reinforcement learning [7], can be modeled using a generalized form of coupled replicator equations.

While replicator dynamics was introduced originally for evolutionary game theory [8], the relationship between reinforcement learning and replicator equations has been developed only recently [9]. Here, we extend these considerations to multiagent systems, introducing the theory behind a previously reported game-theoretic model [10]. We show that replicator dynamics emerges as a special case of the continuous-time limit for multiagent reinforcement-learning systems. The overall approach, though, establishes a general framework for dynamical-systems analyses of adaptive behavior in collectives.

Notably, in learning with perfect memory, our model reduces to the form of a multipopulation replicator equation introduced in Ref. [11]. For two agents with perfect memory interacting via a zero-sum rock-scissors-paper game, the dynamics exhibits Hamiltonian chaos [10]. In contrast, as we show here, with memory decay, multiagent systems generally become dissipative and display the full range of nonlinear dynamical behaviors, including limit cycles, intermittency, and deterministic chaos.

Our multiagent model begins with simple reinforcement-learning agents. To clarify the development, we assume that there are two such agents X and Y that at each time step take one of N actions: $i=1,\ldots,N$. Let the probability for X to choose action i be $x_i(n)$ and $y_i(n)$ for Y, where n is the number of the learning iterations from the initial state at $n=0$. Agents' choice distributions at time n are $\mathbf{x}(n)=(x_1(n),\ldots,x_N(n))$ and $\mathbf{y}(n)=(y_1(n),\ldots,y_N(n))$, with $\sum_i x_i(n)=\sum_i y_i(n)=1$.

Let R_{ij}^X and R_{ij}^Y denote the rewards for X taking action i and Y action j at step n, respectively. Given these actions, X's and Y's memories, $Q_i^X(n)$ and $Q_j^Y(n)$, of the past benefits from their actions are governed by

$$Q_i^X(n+1)-Q_i^X(n)=R_{ij}^X-\alpha_X Q_i^X(n) \text{ and}$$
$$Q_j^Y(n+1)-Q_j^Y(n)=R_{ij}^Y-\alpha_Y Q_j^Y(n), \quad (1)$$

where $\alpha_X,\alpha_Y\in[0,1)$ control each agent's memory decay rate and $Q_i^X(0)=Q_j^Y(0)=0$. The agents choose their next actions according to the Q's, updating their choice distributions as follows:

$$x_i(n)=\frac{e^{\beta_X Q_i^X(n)}}{\sum_j e^{\beta_X Q_j^X(n)}} \text{ and } y_j(n)=\frac{e^{\beta_Y Q_j^Y(n)}}{\sum_j e^{\beta_Y Q_j^Y(n)}}, \quad (2)$$

where $\beta_X,\beta_Y\in[0,\infty]$ control the learning sensitivity: how much the current choice distributions are affected by past rewards. Using Eq. (2), the dynamic governing the change in agent state is given by

$$x_i(n+1)=\frac{x_i(n)e^{\beta_X Q_i^X(n+1)-Q_i^X(n)}}{\sum_j x_j(n)e^{\beta_X Q_j^X(n+1)-Q_j^X(n)}}, \quad (3)$$

and similarly for $y_j(n+1)$.

Consider the continuous-time limit corresponding to agents performing a large number of actions [iterates of Eqs. (1)] for each choice distribution update [iterates of Eq. (3)]. In this case, we have two different time scales—that for agent-agent interactions and that for learning. We assume that the learning dynamics is very slow compared to interactions and so $\mathbf{x}$ and $\mathbf{y}$ are essentially constant during the latter. Then, based on Eq. (3), continuous-time learning for agent X is governed by

*Electronic address: yusto@bdc.brain.riken.go.jp
†Electronic address: chaos@santafe.edu

1063-651X/2003/67(1)/015206(4)$20.00 67 015206-1

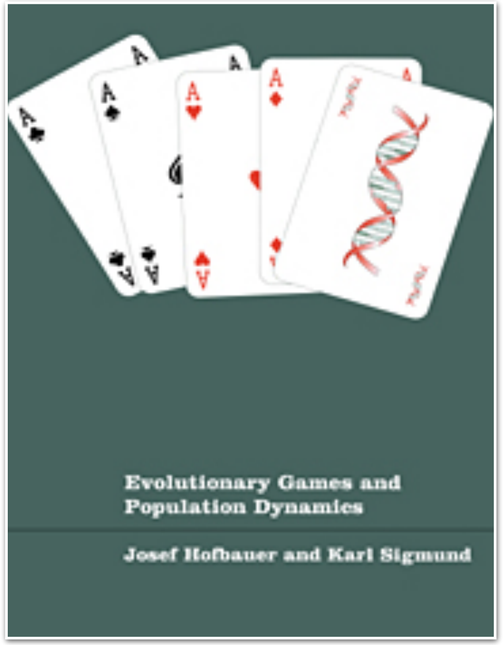
Sato & Crutchfield, 2003

The challenge

Disciplinary divides a clustering attempt

Biology

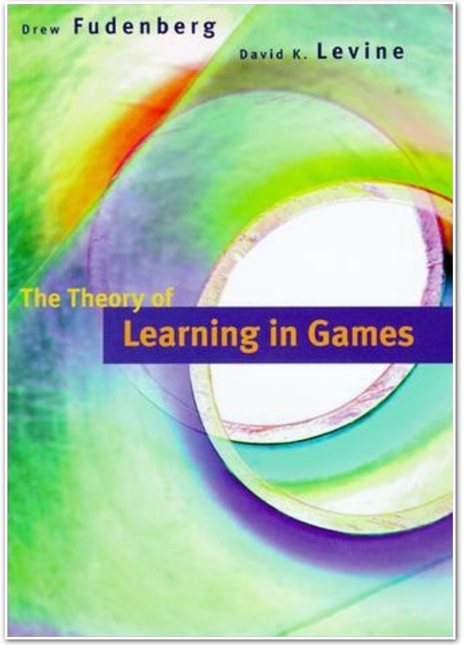
Evolutionary Game Theory



Hofbauer & Sigmund, 1998

Economics

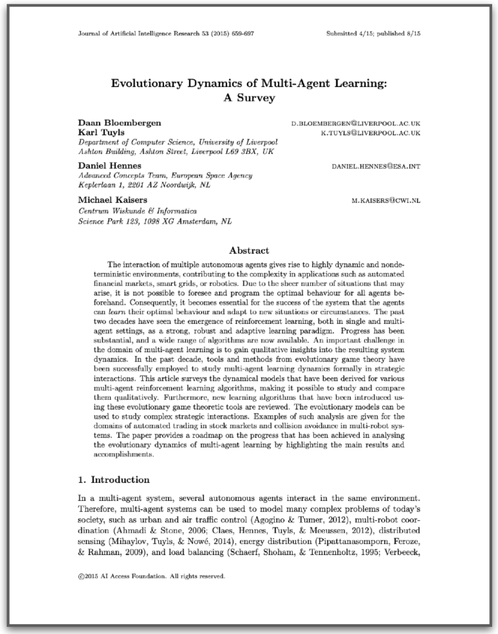
Learning in Games



Fudenberg & Levine, 1998

Computer Science

Machine Learning Dynamics



Journal of Artificial Intelligence Research 53 (2015) 659-697 Submitted 4/15; published 8/15

Evolutionary Dynamics of Multi-Agent Learning: A Survey

Daan Bloembergen D.BLOEMBERGEN@LIVERPOOL.AC.UK
Karl Tuyls K.TUYLS@LIVERPOOL.AC.UK
Department of Computer Science, University of Liverpool
Ashton Building, Ashton Street, Liverpool L69 3BX, UK
Daniel Hennes DANIEL.HENNES@ESA.INT
Advanced Concepts Team, European Space Agency
Keplerlaan 1, 2201 AZ Noordwijk, NL
Michael Kaisers M.KAISERS@CWI.NL
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, NL

Abstract

The interaction of multiple autonomous agents gives rise to highly dynamic and nondeterministic environments, contributing to the complexity in applications such as automated financial markets, smart grids, or robotics. Due to the sheer number of situations that may arise, it is not possible to foresee and program the optimal behaviour for all agents beforehand. Consequently, it becomes essential for the success of the system that the agents can learn their optimal behaviour and adapt to new situations or circumstances. The past two decades have seen the emergence of reinforcement learning, both in single and multi-agent settings, as a strong, robust and adaptive learning paradigm. Progress has been substantial, and a wide range of algorithms are now available. An important challenge in the domain of multi-agent learning is to gain qualitative insights into the resulting system dynamics. In the past decade, tools and methods from evolutionary game theory have been successfully employed to study multi-agent learning dynamics formally in strategic interactions. This article surveys the dynamical models that have been derived for various multi-agent reinforcement learning algorithms, making it possible to study and compare them qualitatively. Furthermore, new learning algorithms that have been introduced using these evolutionary game theoretic tools are reviewed. The evolutionary models can be used to study complex strategic interactions. Examples of such analysis are given for the domains of automated trading in stock markets and collision avoidance in multi-robot systems. The paper provides a roadmap on the progress that has been achieved in analysing the evolutionary dynamics of multi-agent learning by highlighting the main results and accomplishments.

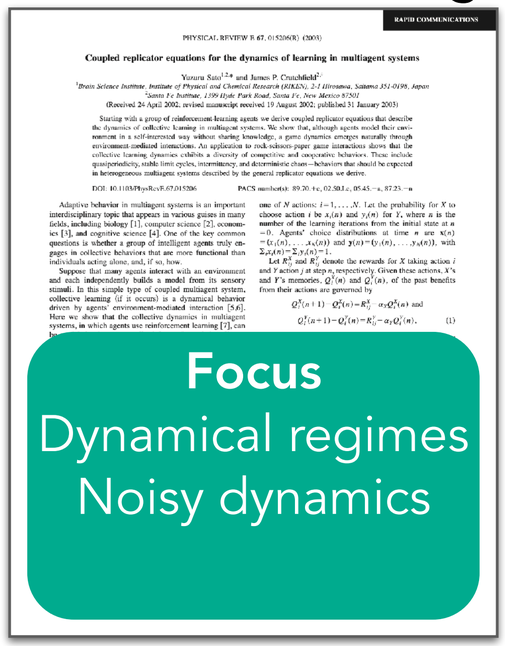
1. Introduction

In a multi-agent system, several autonomous agents interact in the same environment. Therefore, multi-agent systems can be used to model many complex problems of today's society, such as urban and air traffic control (Agogino & Tumer, 2012), multi-robot coordination (Ahmadi & Stone, 2006; Claes, Hennes, Tuyls, & Meeussen, 2012), distributed sensing (Mihaylov, Tuyls, & Nowé, 2014), energy distribution (Pipattanasomporn, Feroze, & Rahman, 2009), and load balancing (Schaerf, Shoham, & Tennenholtz, 1995; Verbeeck,

Bloembergen et al., 2015

Physics

Statistical mechanics learning



RAPID COMMUNICATIONS

PHYSICAL REVIEW E 67, 015206(R) (2003)

Coupled replicator equations for the dynamics of learning in multiagent systems

Yuzuru Sato[1,2,*] and James P. Crutchfield[2]
[1]*Brain Science Institute, Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Saitama 351-0198, Japan*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*
(Received 24 April 2002; revised manuscript received 19 August 2002; published 31 January 2003)

Starting with a group of reinforcement-learning agents we derive coupled replicator equations that describe the dynamics of collective learning in multiagent systems. We show that, although agents model their environment in a self-interested way without sharing knowledge, a game dynamics emerges naturally through environment-mediated interactions. An application to rock-scissors-paper game interactions shows that the collective learning dynamics exhibits a diversity of competitive and cooperative behaviors. These include quasiperiodicity, stable limit cycles, intermittency, and deterministic chaos—behaviors that should be expected in heterogeneous multiagent systems described by the general replicator equations we derive.

DOI: 10.1103/PhysRevE.67.015206 PACS number(s): 89.20.+e, 02.50.Le, 05.45.−a, 87.23.−n

Adaptive behavior in multiagent systems is an important interdisciplinary topic that appears in various guises in many fields, including biology [1], computer science [2], economics [3], and cognitive science [4]. One of the key common questions is whether a group of intelligent agents truly engages in collective behaviors that are more functional than individuals acting alone, and, if so, how.

Suppose that many agents interact with an environment and each independently builds a model from its sensory stimuli. In this simple type of coupled multiagent system, collective learning (if it occurs) is a dynamical behavior driven by agents' environment-mediated interaction [5,6]. Here we show that the collective dynamics in multiagent systems, in which agents use reinforcement learning [7], can be

one of N actions: $i=1,\ldots,N$. Let the probability for X to choose action i be $x_i(n)$ and $y_i(n)$ for Y, where n is the number of the learning iterations from the initial state at $n=0$. Agents' choice distributions at time n are $\mathbf{x}(n)=(x_1(n),\ldots,x_N(n))$ and $\mathbf{y}(n)=(y_1(n),\ldots,y_N(n))$, with $\sum_i x_i(n)=\sum_i y_i(n)=1$.

Let R_{ij}^X and R_{ij}^Y denote the rewards for X taking action i and Y action j at step n, respectively. Given these actions, X's and Y's memories, $Q_i^X(n)$ and $Q_j^Y(n)$, of the past benefits from their actions are governed by

$$Q_i^X(n+1)-Q_i^X(n)=R_{ij}^X-\alpha_X Q_i^X(n) \text{ and}$$
$$Q_j^Y(n+1)-Q_j^Y(n)=R_{ij}^Y-\alpha_Y Q_j^Y(n). \quad (1)$$

Focus
Dynamical regimes
Noisy dynamics

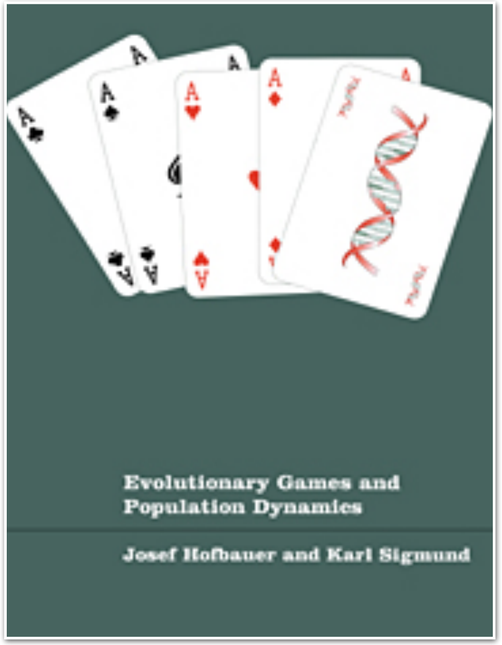
Sato & Crutchfield, 2003

The challenge

Disciplinary divides a clustering attempt

Biology

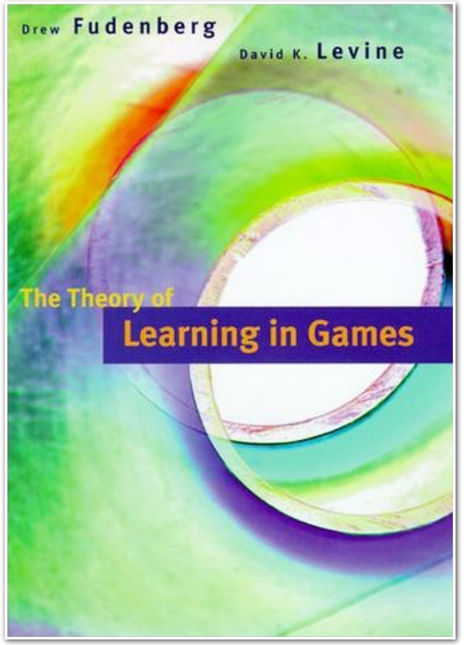
Evolutionary Game Theory



Hofbauer & Sigmund, 1998

Economics

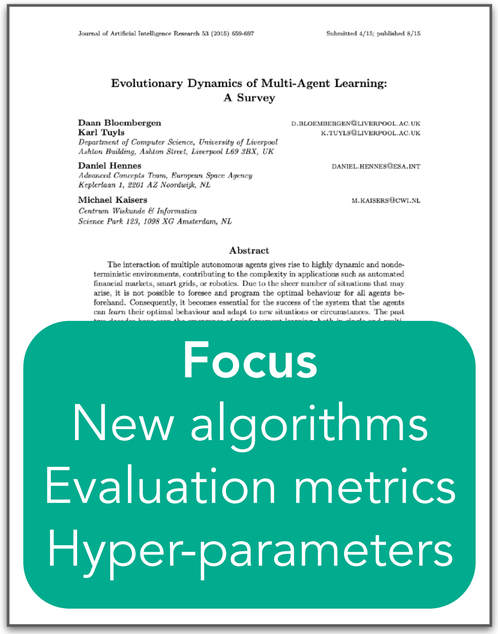
Learning in Games



Fudenberg & Levine, 1998

Computer Science

Machine Learning Dynamics



Journal of Artificial Intelligence Research 53 (2015) 659-697 Submitted 4/15; published 8/15

Evolutionary Dynamics of Multi-Agent Learning: A Survey

Daan Bloembergen D.BLOEMBERGEN@LIVERPOOL.AC.UK
Karl Tuyls K.TUYLS@LIVERPOOL.AC.UK
Department of Computer Science, University of Liverpool
Ashton Building, Ashton Street, Liverpool L69 3BX, UK

Daniel Hennes DANIEL.HENNES@ESA.INT
Advanced Concepts Team, European Space Agency
Keplerlaan 1, 2201 AZ Noordwijk, NL

Michael Kaisers M.KAISERS@CWI.NL
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, NL

Abstract

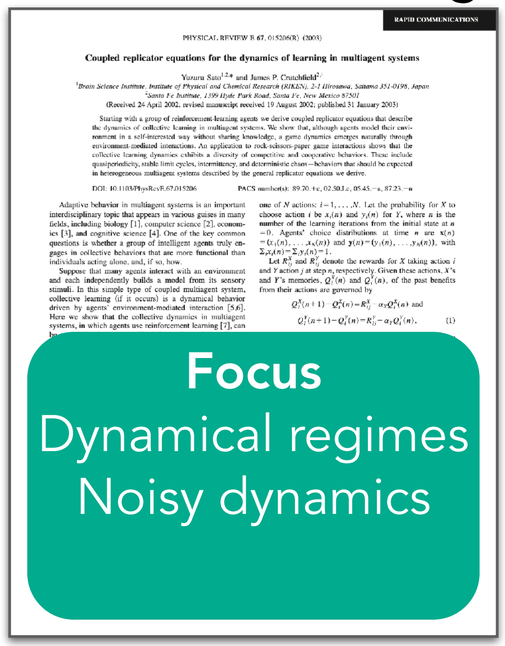
The interaction of multiple autonomous agents gives rise to highly dynamic and nondeterministic environments, contributing to the complexity in applications such as automated financial markets, smart grids, or robotics. Due to the sheer number of situations that may arise, it is not possible to foresee and program the optimal behaviour for all agents beforehand. Consequently, it becomes essential for the success of the system that the agents can learn their optimal behaviour and adapt to new situations or circumstances. The past

Focus
New algorithms
Evaluation metrics
Hyper-parameters

Bloembergen et al., 2015

Physics

Statistical mechanics learning



RAPID COMMUNICATIONS

PHYSICAL REVIEW E 67, 015206(R) (2003)

Coupled replicator equations for the dynamics of learning in multiagent systems

Yuzuru Sato[1,2,*] and James P. Crutchfield[2]
[1]*Brain Science Institute, Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Saitama 351-0198, Japan*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*
(Received 24 April 2002; revised manuscript received 19 August 2002; published 31 January 2003)

Starting with a group of reinforcement-learning agents we derive coupled replicator equations that describe the dynamics of collective learning in multiagent systems. We show that, although agents model their environment in a self-interested way without sharing knowledge, a game dynamics emerges naturally through environment-mediated interactions. An application to rock-scissors-paper game interactions shows that the collective learning dynamics exhibits a diversity of competitive and cooperative behaviors. These include quasiperiodicity, stable limit cycles, intermittency, and deterministic chaos—behaviors that should be expected in heterogeneous multiagent systems described by the general replicator equations we derive.

DOI: 10.1103/PhysRevE.67.015206 PACS number(s): 89.20.+e, 02.50.Le, 05.45.−a, 87.23.−n

Adaptive behavior in multiagent systems is an important interdisciplinary topic that appears in various guises in many fields, including biology [1], computer science [2], economics [3], and cognitive science [4]. One of the key common questions is whether a group of intelligent agents truly engages in collective behaviors that are more functional than individuals acting alone, and, if so, how.

Suppose that many agents interact with an environment and each independently builds a model from its sensory stimuli. In this simple type of coupled multiagent system, collective learning (if it occurs) is a dynamical behavior driven by agents' environment-mediated interaction [5,6]. Here we show that the collective dynamics in multiagent systems, in which agents use reinforcement learning [7], can

one of N actions: $i=1,\ldots,N$. Let the probability for X to choose action i be $x_i(n)$ and $y_i(n)$ for Y, where n is the number of the learning iterations from the initial state at $n=0$. Agents' choice distributions at time n are $\mathbf{x}(n)=(x_1(n),\ldots,x_N(n))$ and $\mathbf{y}(n)=(y_1(n),\ldots,y_N(n))$, with $\sum_i x_i(n)=\sum_i y_i(n)=1$.

Let R_{ij}^X and R_{ij}^Y denote the rewards for X taking action i and Y action j at step n, respectively. Given these actions, X's and Y's memories, $Q_i^X(n)$ and $Q_j^Y(n)$, of the past benefits from their actions are governed by

$$Q_i^X(n+1)-Q_i^X(n)=R_{ij}^X-\alpha_X Q_i^X(n) \text{ and}$$
$$Q_j^Y(n+1)-Q_j^Y(n)=R_{ij}^Y-\alpha_Y Q_j^Y(n). \quad (1)$$

Focus
Dynamical regimes
Noisy dynamics

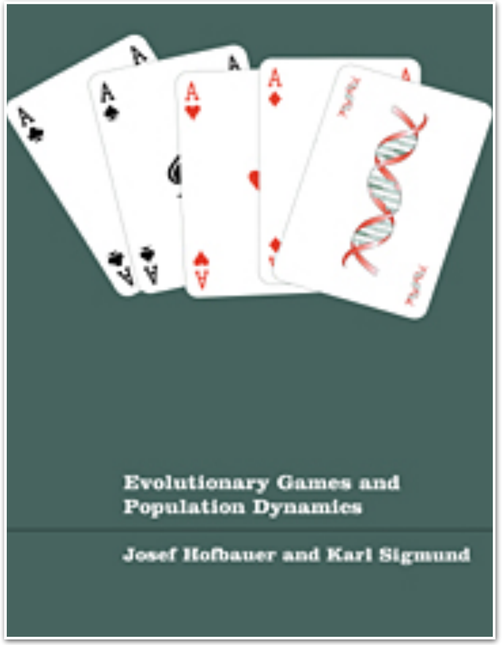
Sato & Crutchfield, 2003

The challenge

Disciplinary divides a clustering attempt

Biology

Evolutionary Game Theory



Hofbauer & Sigmund, 1998

Economics

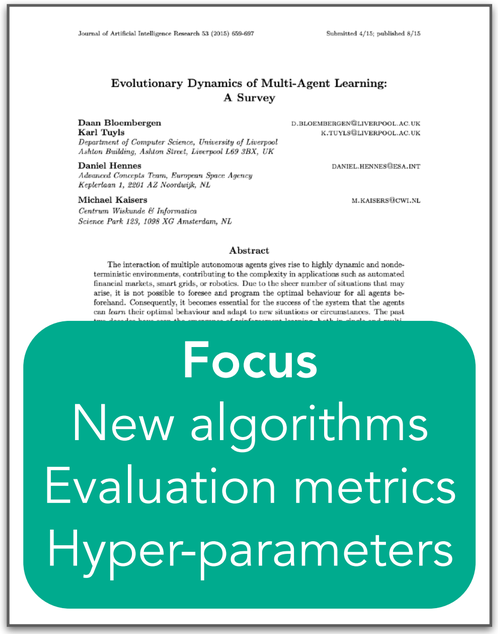
Learning in Games

Focus
Foundations for game equilibria

Fudenberg & Levine, 1998

Computer Science

Machine Learning Dynamics

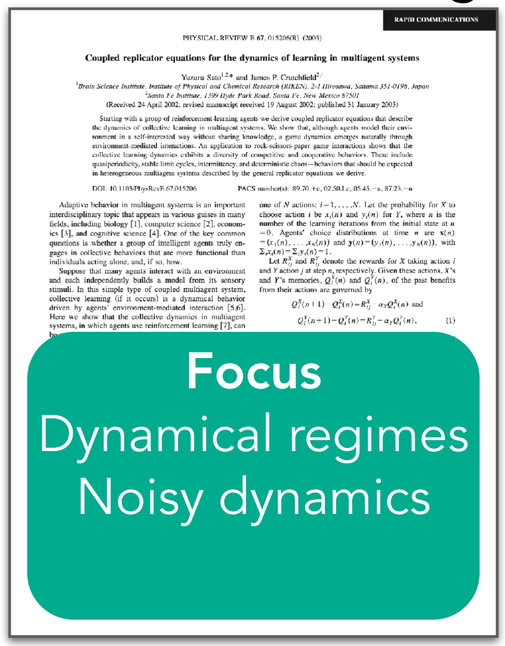


Focus
New algorithms
Evaluation metrics
Hyper-parameters

Bloembergen et al., 2015

Physics

Statistical mechanics learning



Focus
Dynamical regimes
Noisy dynamics

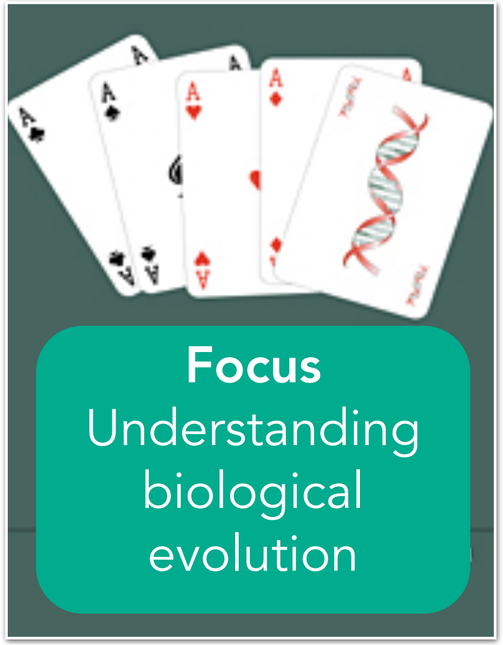
Sato & Crutchfield, 2003

The challenge

Disciplinary divides a clustering attempt

Biology

Evolutionary Game Theory



Focus
Understanding biological evolution

Hofbauer & Sigmund, 1998

Economics

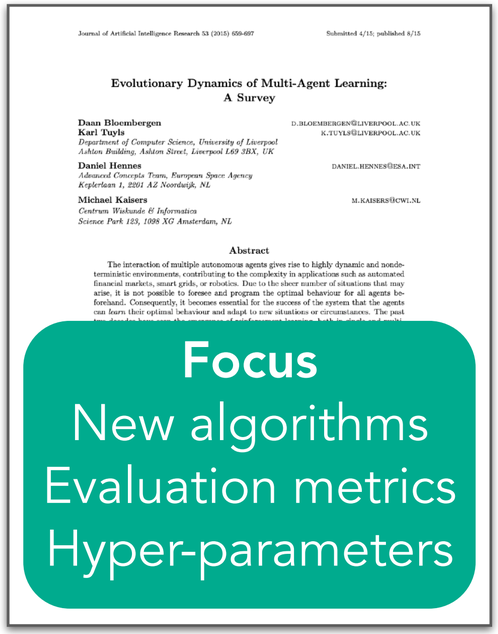
Learning in Games

Focus
Foundations for game equilibria

Fudenberg & Levine, 1998

Computer Science

Machine Learning Dynamics

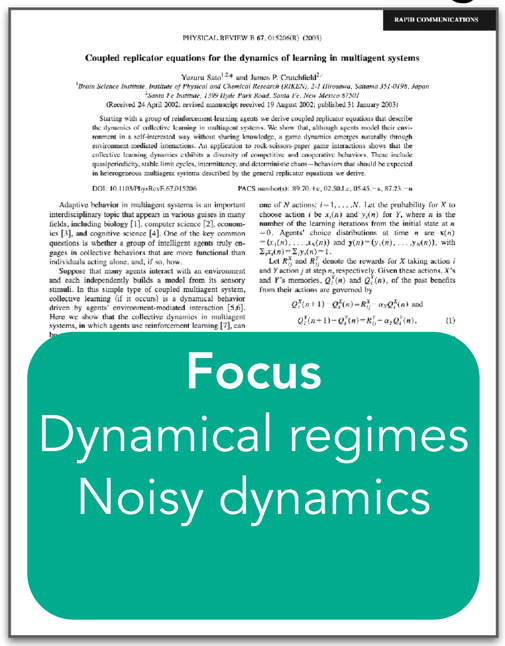


Focus
New algorithms
Evaluation metrics
Hyper-parameters

Bloembergen et al., 2015

Physics

Statistical mechanics learning



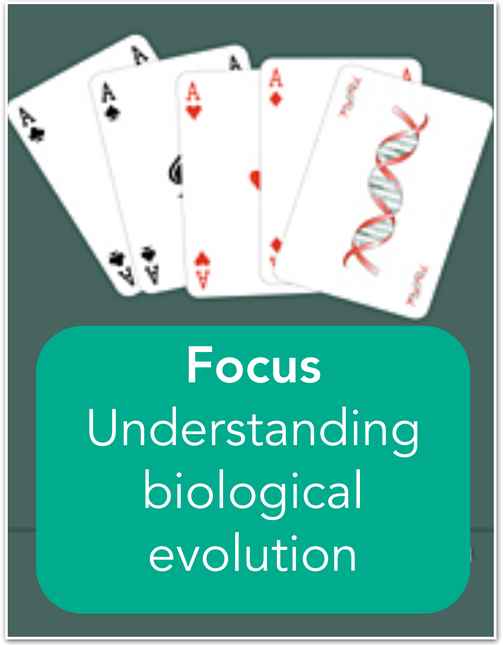
Focus
Dynamical regimes
Noisy dynamics

Sato & Crutchfield, 2003

The challenge

Disciplinary divides a clustering attempt

Biology
Evolutionary Game Theory



Focus
Understanding biological evolution

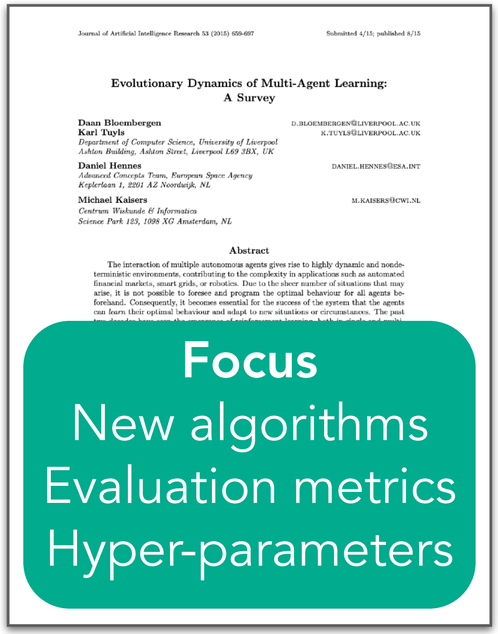
Hofbauer & Sigmund, 1998

Economics
Learning in Games

Focus
Foundations for game equilibria

Fudenberg & Levine, 1998

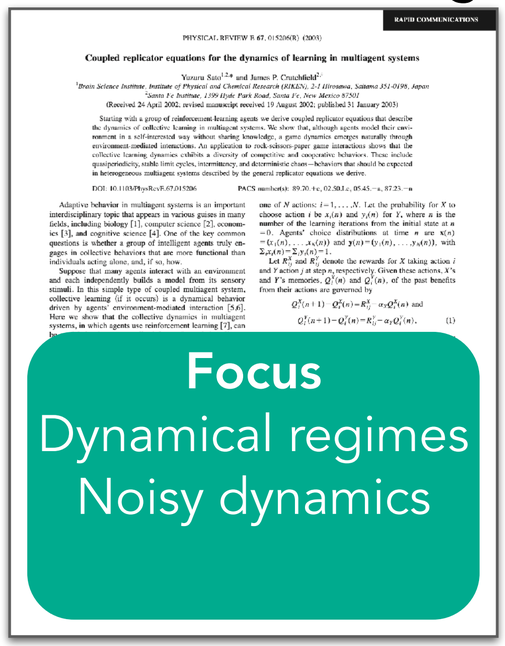
Computer Science
Machine Learning Dynamics



Focus
New algorithms
Evaluation metrics
Hyper-parameters

Bloembergen et al., 2015

Physics
Statistical mechanics learning



Focus
Dynamical regimes
Noisy dynamics

Sato & Crutchfield, 2003

➡ Towards a unified treatment

Refining Environments

What do we want? ➡ a DynGym - a gym for learning *dynamics* ?

Refining Environments

What do we want? ➡ a DynGym - a gym for learning *dynamics* ?

- **(Stylised) single agent capabilities** (~Osband et al., 2019):
  exploration, memory, credit assignment,…

Refining Environments

What do we want? ➡ a DynGym - a gym for learning *dynamics* ?

- **(Stylised) single agent capabilities** (~Osband et al., 2019):
  exploration, memory, credit assignment,…
- **(Stylised) multi-agent capabilities**:
  global exploration, coordination, cooperation (trust), competition (not exploitable),…

Refining Environments

What do we want? ➡ a DynGym - a gym for learning *dynamics* ?

- **(Stylised) single agent capabilities** (~Osband et al., 2019):
  exploration, memory, credit assignment,…
- **(Stylised) multi-agent capabilities**:
  global exploration, coordination, cooperation (trust), competition (not exploitable),…
- **(Stylised) applications**:
  AI safety (~Leike et al. 2017), AI governance (~The Anh Han et al. 2020),
  Sustainability in social-ecological dilemmas (~Barfuss et al. 2020), …

Refining Environments

What do we want? ➡ a DynGym - a gym for learning *dynamics* ?

- **(Stylised) single agent capabilities** (~Osband et al., 2019):
  exploration, memory, credit assignment,…
- **(Stylised) multi-agent capabilities**:
  global exploration, coordination, cooperation (trust), competition (not exploitable),…
- **(Stylised) applications**:
  AI safety (~Leike et al. 2017), AI governance (~The Anh Han et al. 2020),
  Sustainability in social-ecological dilemmas (~Barfuss et al. 2020), …

➡ Benchmarks?

Refining Agents

Advancing agents / relaxing assumptions

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**
- RRL agents with **stylised function approximations**

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**
- RRL agents with **stylised function approximations**
- RRL agents with **finite memory batch** > noisy dynamics

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**
- RRL agents with **stylised function approximations**
- RRL agents with **finite memory batch** > noisy dynamics
- RRL agents with **different learning paradigms** (imitation, belief, …)

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**
- RRL agents with **stylised function approximations**
- RRL agents with **finite memory batch** > noisy dynamics
- RRL agents with **different learning paradigms** (imitation, belief, …)
- …

Refining Agents

Advancing agents / relaxing assumptions

- RRL agents with **partial observability**
- RRL agents with **stylised function approximations**
- RRL agents with **finite memory batch** > noisy dynamics
- RRL agents with **different learning paradigms** (imitation, belief, …)
- …

➡ **Taxonomy?**

Refining Evaluations & Analysis

What are we looking for?

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

➡ **Principles** for their dependence
on agents' hyper-parameters and environment characteristics

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

  ➡ **Principles** for their dependence
  on agents' hyper-parameters and environment characteristics

- What else?

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

  ➡ **Principles** for their dependence
  on agents' hyper-parameters and environment characteristics

- What else?

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

  ➡ **Principles** for their dependence
  on agents' hyper-parameters and environment characteristics

- What else?

- ➡ Complex environments <=> empirical (stochastic) game theoretic analysis (Tuyls et al., 2018)

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

  ➡ **Principles** for their dependence
  on agents' hyper-parameters and environment characteristics

- What else?

- ➡ Complex environments <=> empirical (stochastic) game theoretic analysis (Tuyls et al., 2018)
- ➡ Large-agent limit <=> Mean-field games (Elie et al., 2020)

Refining Evaluations & Analysis

What are we looking for?

- Convergence to equilibria -> multi-stability -> other dynamical regimes (periodic orbits, chaos)

  ➡ **Principles** for their dependence
  on agents' hyper-parameters and environment characteristics

- What else?

- ➡ Complex environments <=> empirical (stochastic) game theoretic analysis (Tuyls et al., 2018)
- ➡ Large-agent limit <=> Mean-field games (Elie et al., 2020)
- ➡ … other connections?

Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges

Conclusions

Collective Learning Dynamics - a useful semi-formal method
for helping with MARL's challenges **=> unified treatment**

Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges **=> unified treatment**

↑

Applications

> Problem-Method Fit?

Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges **=> unified treatment**

Applications

> Problem-Method Fit?

Refining Environments

> Benchmarks?

Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges **=> unified treatment**

Applications

> Problem-Method Fit?

Refining Agents

> Taxonomy?

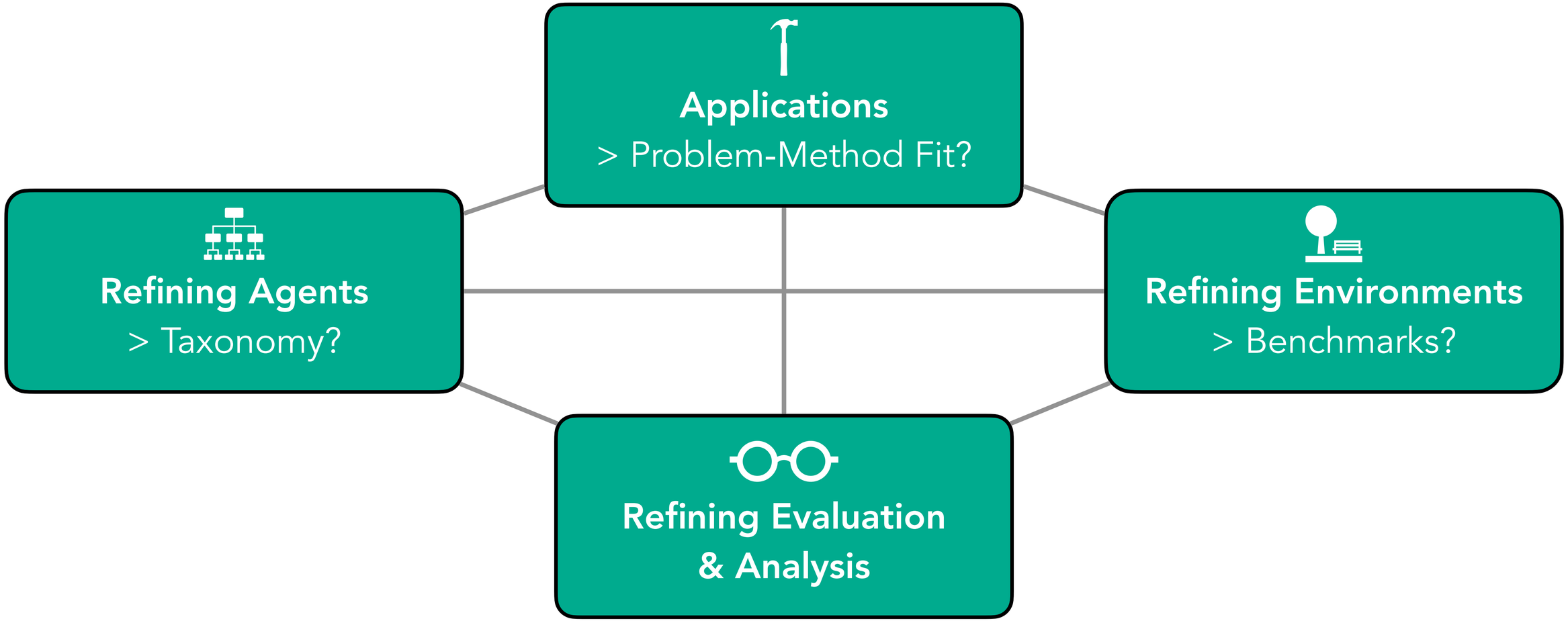
Refining Environments

> Benchmarks?

Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges **=> unified treatment**



Conclusions

Collective Learning Dynamics - a useful semi-formal method

for helping with MARL's challenges **=> unified treatment**

Applications
> Problem-Method Fit?

Refining Agents
> Taxonomy?

Thank you

Refining Environments
> Benchmarks?

Refining Evaluation & Analysis

Please get in touch
wolfram.barfuss@gmail.com
wbarfuss.github.io

References

- Barfuss, W., Donges, J.F., Vasconcelos, V.V., Kurths, J. and Levin, S.A., 2020. Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. Proceedings of the National Academy of Sciences, 117(23), pp.12915-12922.
- Bloembergen, D., Tuyls, K., Hennes, D. and Kaisers, M., 2015. Evolutionary dynamics of multi-agent learning: A survey. Journal of Artificial Intelligence Research, 53, pp.659-697.
- Elie, R., Pérolat, J., Laurière, M., Geist, M. and Pietquin, O., 2020, April. On the convergence of model free learning in mean field games. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 7143-7150).
- Fudenberg, D., Drew, F., Levine, D.K. and Levine, D.K., 1998. The theory of learning in games (Vol. 2). MIT press.
- Han, T.A., Pereira, L.M., Santos, F.C. and Lenaerts, T., 2020. To Regulate or Not: A Social Dynamics Analysis of an Idealised AI Race. Journal of Artificial Intelligence Research, 69, pp.881-921.
- Hofbauer, J. and Sigmund, K., 1998. Evolutionary games and population dynamics. Cambridge university press.
- Leike, J., Martic, M., Krakovna, V., Ortega, P.A., Everitt, T., Lefrancq, A., Orseau, L. and Legg, S., 2017. AI safety gridworlds. arXiv preprint arXiv:1711.09883.
- Osband, I., Doron, Y., Hessel, M., Aslanides, J., Sezener, E., Saraiva, A., McKinney, K., Lattimore, T., Szepesvari, C., Singh, S. and Van Roy, B., 2019. Behaviour suite for reinforcement learning. arXiv preprint arXiv:1908.03568.
- Pérolat, J., Leibo, J.Z., Zambaldi, V.F., Beattie, C., Tuyls, K. and Graepel, T., 2017, January. A multi-agent reinforcement learning model of common-pool resource appropriation. In NIPS.
- Sato, Y. and Crutchfield, J.P., 2003. Coupled replicator equations for the dynamics of learning in multiagent systems. Physical Review E, 67(1), p.015206.
- Tuyls, K., Perolat, J., Lanctot, M., Leibo, J.Z. and Graepel, T., 2018. A generalised method for empirical game theoretic analysis. arXiv preprint arXiv:1803.06376.