

# Multi-agent learning dynamics

Citation for published version (APA):

Bloembergen, D. (2015). *Multi-agent learning dynamics*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20150521db>

## Document status and date:

Published: 01/01/2015

## DOI:

[10.26481/dis.20150521db](https://doi.org/10.26481/dis.20150521db)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Summary

Within the field of Artificial Intelligence, autonomous decision makers are referred to as *agents*. When multiple such decision makers interact within a common environment we are talking about a multi-agent system. Multi-agent systems are well-suited to model a variety of complex large-scale problems in our society, such as automated financial markets, the smart grid, (air) traffic control, and multi-robot systems. Each individual agent affects its environment by taking decisions and executing actions, thereby influencing the actions taken by other agents as well, thus yielding a highly dynamic environment. Due to the sheer number of situations that may arise it is not possible to foresee and program the optimal behaviour for each one beforehand. Consequently, the success of the system depends on the ability of the agents to *learn* to optimise their behaviour and adapt to new situations or circumstances. The field of *multi-agent learning* is involved with precisely this problem.

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, as the complexity of multi-agent interactions renders theoretical analysis difficult. In recent years, *evolutionary game theory* has been taken up as a promising paradigm within which to study multi-agent learning formally. Its promise rests on the proven link between the behaviour of simple reinforcement learning algorithms, such as learning automata, and the *replicator dynamics* of evolutionary game theory. The replicator dynamics predict the expected behaviour of reinforcement learning agents in normal-form games.

In this dissertation we investigate the dynamics of learning in multi-agent systems by building on the evolutionary game theoretic approach to multi-agent learning. The thesis comprises four main contributions, which are briefly summarised here.

First, in Chapter 3 we survey current research on the evolutionary dynamics of multi-agent learning, summarising the main results in this field so far. The formal relation between the replicator dynamics and multi-agent reinforcement learning is detailed and extensions of the model to various learning algorithms are presented, both with discrete and continuous action spaces and in both single-state and multi-state environments. We show how this analysis leads to new insights that can bridge two seemingly diverse streams of research.

Secondly, in Chapter 4 we show how lenience can enable cooperation in multi-

agent systems. A lenient agent ignores low rewards resulting from mis-coordination due to the fact that the other agents are also learning. We propose *lenient frequency-adjusted Q-learning*, investigate the dynamics of this new algorithm, and compare its convergence properties to several other state-of-the-art learning algorithms. Experiments in normal-form games and on graphs show that lenience improves coordination in the well-known strategic scenario given by the stag hunt game.

Thirdly, in Chapter 5 we propose a new dynamical model for the *continuous action iterated prisoner's dilemma* on graphs, based on methods and techniques from control theory. This model allows us to analytically study the evolution of cooperation on arbitrary complex networks. We show that allowing a continuous rather than discrete action space improves cooperation. Moreover, we investigate the influence of structural network properties on the final outcome. In particular we compare the small-world and scale-free network model, both of which exhibit properties that are typically found in real-world social or technological networks. We then extend the model to allow for external influence by which the network dynamics can be controlled and we propose an iterative control algorithm that can drive the network to any desired state while minimising the control effort.

Finally, in Chapter 6 we apply the evolutionary dynamical analysis to the complex domain of automated trading in stock markets. By focusing on high-level trading strategies rather than atomic actions we can use the replicator dynamics to analyse the evolutionary strength of each strategy. In particular, we investigate the value of information in stock markets under influence of both noise and cost, and find that more information is not always better, in particular when that information is expensive or uncertain. Moreover, we compare two trading strategies that are common to real stock markets – chartists and fundamentalists – and investigate under which settings both can be sustained in an evolutionary market.

The findings presented in this dissertation contribute to a better understanding of the dynamics of learning in multi-agent systems. We have surveyed recent advances in the intersection between evolutionary game theory and multi-agent learning, and have subsequently used these methods to study cooperation through lenience in multi-agent systems. We then extended the evolutionary framework to networked interactions and investigated the evolution of cooperation in societies of self-interested decision makers. Finally, we have applied the evolutionary framework to the study and analysis of trading in stock markets, thereby showing the applicability of the framework to complex multi-agent interactions and thus proving its value beyond the analysis of stylised games.

# Samenvatting

Binnen het vakgebied kunstmatige intelligentie worden eenheden die zelfstandig beslissingen kunnen nemen *agenten* genoemd. Wanneer meerdere agenten met elkaar in interactie zijn binnen een gezamenlijke omgeving, spreken we van een *multi-agent systeem*. Multi-agent systemen zijn uitermate geschikt om een scala aan complexe problemen uit onze huidige maatschappij te modelleren. Denk bijvoorbeeld aan geautomatiseerde financiële markten, slimme energienetwerken, en het regelen van (lucht)verkeer. Elke individuele agent beïnvloedt zijn omgeving door het nemen van beslissingen en het uitvoeren van acties, waardoor indirect ook de acties van andere agenten beïnvloed worden. Dit leidt tot een zeer dynamische omgeving. Het schier oneindig aantal situaties dat zo kan optreden maakt het onmogelijk om vooraf precies te bepalen welk gedrag op elk moment optimaal is. Het is hierom van groot belang dat agenten in staat zijn om te *leren* en zich aan te passen aan nieuwe situaties en omstandigheden. Het onderzoeksgebied *multi-agent learning* houdt zich hiermee bezig.

*Reinforcement learning* heeft zich in de laatste twee decennia bewezen als een robuuste en adaptieve leer methode voor multi-agent systemen. Er is substantiële vooruitgang geboekt, en een verscheidenheid aan leeralgoritmen is nu beschikbaar. Een belangrijke uitdaging binnen het gebied van multi-agent learning is om kwalitatief en theoretisch inzicht te krijgen in de dynamiek van het leerproces, wat bemoeilijkt wordt door de complexiteit die de interactie in multi-agent systemen met zich meebrengt. *Evolutionaire speltheorie* biedt uitkomst, en is recentelijk omarmt als een veelbelovend raamwerk waarbinnen multi-agent learning formeel bestudeerd kan worden. Deze belofte rust op de bewezen connectie tussen het gedrag van eenvoudige leeralgoritmen, zoals leerautomaten, en de *replicator dynamics* van de evolutionaire speltheorie. De replicator dynamics voorspellen het verwachte gedrag van lerende agenten in strategische interacties die beschreven zijn in zogeheten *normal-form games*.

In dit proefschrift onderzoeken we de leerdynamiek van multi-agent systemen door voort te borduren op het hiervoor beschreven evolutionair speltheoretisch raamwerk. Het proefschrift omvat vier kernbijdragen die we hier kort samenvatten.

Allereerst geven we in Hoofdstuk 3 een overzicht van het evolutionair speltheoretisch raamwerk voor multi-agent learning, waarbij we de belangrijkste conclusies die tot dusver zijn getrokken samenvatten. We beschrijven de formele relatie tussen de replicator dynamics en leerautomaten in normal-form games, almede extensies van de theorie naar complexere leeralgoritmen en omgevingen. We laten zien hoe een dergelijke analyse tot nieuwe inzichten kan leiden waarbij twee ogenschijnlijk diverse

groepen leeralgoritmen met elkaar verbonden worden.

Ten tweede laten we in Hoofdstuk 4 zien hoe *lenience*, vergelijkbaar met welwillendheid, samenwerking mogelijk maakt in multi-agent systemen. Een *lenient* agent negeert slechte resultaten als gevolg van miscoördinatie die enkel voortkomt uit het feit dat andere agenten eveneens nog aan het leren zijn. We presenteren *lenient frequency-adjusted Q-learning*, onderzoeken de leerdynamica van dit nieuwe algoritme, en vergelijken de convergentie-eigenschappen met enkele andere recente leeralgoritmen. Experimenten in normal-form games en op grafen laten zien dat *lenience* inderdaad coördinatie verbetert in het bekende strategische scenario van de *stag hunt*.

Ten derde presenteren we in Hoofdstuk 5 een nieuw dynamisch model voor het *continuous-action iterated prisoner's dilemma* op grafen, gebaseerd op methodes uit het onderzoeksgebied van de meet- en regeltechniek. Dit model stelt ons in staat om de evolutie van samenwerking in complexe netwerken te analyseren. We laten zien dat de continue actieruimte van dit model samenwerking bevordert ten opzichte van een discrete actieruimte. Daarnaast onderzoeken we de invloed van structurele eigenschappen van het netwerk op de resulterende dynamiek. Vervolgens breiden we het model uit om externe aansturing van het netwerk mogelijk te maken door individuele knopen van het netwerk te beïnvloeden. Tevens presenteren we een iteratief algoritme dat in staat is om het netwerk naar een gewenste staat te sturen waarbij tegelijkertijd de vereiste externe invloed wordt geminimaliseerd.

Tot slot passen we in Hoofdstuk 6 de evolutionair-dynamische analyse toe op het complexe domein van geautomatiseerde aandeelhandel. Met behulp van heuristische methodes kunnen we de replicator dynamics gebruiken om de evolutionaire sterkte van complexe handelsstrategieën te analyseren. Specifiek onderzoeken we de waarde van informatie in de aandelenhandel onder invloed van kosten en onzekerheid. We concluderen dat meer informatie niet altijd leidt tot betere resultaten, met name wanneer informatie duur of onzeker is. Daarnaast vergelijken we twee strategieën die veelvoorkomend zijn in echte aandelenmarkten en onderzoeken onder welke omstandigheden beide strategieën kunnen overleven in een evoluerende markt.

De bevindingen die we hebben gepresenteerd in dit proefschrift dragen bij aan een beter begrip van de leerdynamiek van multi-agent systemen. We hebben een overzicht geschetst van het evolutionair speltheoretisch raamwerk voor de studie van multi-agent systemen, en vervolgens deze methodes en technieken toegepast bij het bestuderen van coördinatie door middel van *lenience*. Hierna hebben we het raamwerk uitgebreid naar interacties in grafen, en aan de hand daarvan samenwerking in complexe netwerken bestudeerd. Tot slot hebben we de evolutionaire methode gebruikt voor de analyse van aandeelhandel. Hiermee hebben we aangetoond dat het raamwerk toepasbaar is op complexe multi-agent systemen, en daardoor ook van waarde is buiten gestileerde speltheoretische interacties.