Multi-Agent Learning for Resource Allocation Problems

THÈSE N^{O} 5599 (2013)

PRÉSENTÉE LE 18 JANVIER 2013

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

LABORATOIRE D'INTELLIGENCE ARTIFICIELLE

PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

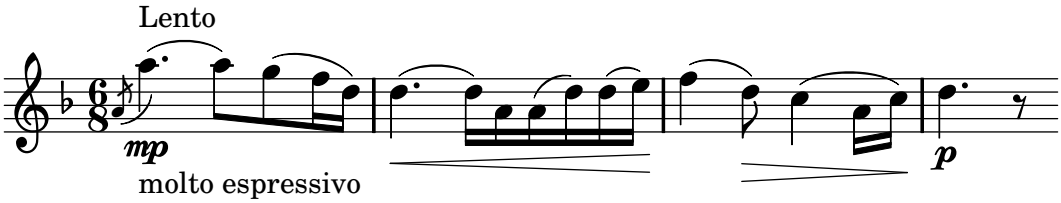
Luděk CIGLER

acceptée sur proposition du jury:

Prof. B. Rimoldi, président du jury
Prof. B. Faltings, directeur de thèse
Prof. J.-P. Hubaux, rapporteur
Prof. K. S. Larson, rapporteur
Prof. D. Parkes, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013



— Antonín Dvořák, String Quartet in F major, “American”

In memory of my grandmother

Acknowledgements

First of all, I would like to express my deepest gratitude to my thesis advisor, Prof. Boi Faltings, for his research supervision and enduring support during my four years at EPFL. I could never have imagined more open-minded and accommodating advisor. I am also extremely thankful to Prof. David Parkes for introducing me to the exciting field of game theory during his seminar at EPFL, and for providing me the opportunity to visit his lab at Harvard. His feedback on my work has been invaluable. Next, I would like to thank Prof. Kate Larson for her detailed comments on my second conference paper. I would also like to extend my acknowledgments to Prof. Jean-Pierre Hubaux for his valuable comments on the possibilities of practical application of this work.

I would have never managed to finish this thesis without the support of my colleagues and friends. My colleagues at the Laboratory of Artificial Intelligence have helped create a friendly and inspiring workplace. I would like to thank them personally, in no particular order: Our two successive secretaries, Marie Decreauzat and Sylvie Thomet, for the grace with which they handled everyday problems, small and big; Brammert Ottens and Goran Radanovic, for being such friendly and understanding officemates; Thomas Léauté, for being an example in research and life; Immanuel Trummer, for being so understanding about my less-than-witty jokes; Florent Garcin, for introducing me to *fondant au chocolat*; Pu Li, for bringing the helicopter; Radek Szymanek, for teaching us about the value of things; Jason Li, for showing what can be accomplished in a short period of time; Christos Dimitrakakis, for being the occasional running buddy; Martin Veselý, for making sure our servers run smoothly; Claudiu Musat, for his help with the Intelligent Agents course, and for making sure *I* make sure our servers run smoothly; Marina Boia, for the excellent marble cake; Ashton Anderson from Stanford University, for showing that research can be fun again; Nicolas Gast, for proving the expected convergence of a generalized coupon collector problem, that has simplified the proofs in this work quite a bit; Jean-Cedric Chappelier and Jamila Sam, for giving me the exciting opportunity to teach in French; and Martin Rajman, for his to-the-point feedback during our lab meetings.

I could not imagine my life in Lausanne without the company of my friends. I would like to thank Petr Sušil, for sharing a taste in Belgian beer; Cristina Ghiurcuta for her enthusiasm in organizing our mountaineering adventures; my flatmates, Albrecht Lindner and Mario Christoudias, for creating a peaceful home; Vlasta Závadová, for her passion for cross-country skiing; Team Dionysos and especially Michel Herren, for keeping me in shape and for bringing their *joie de vivre* in my life; Steve Bettschen, for supplying me with his excellent, inspiration-

inducing wine; Alain Balleret, for always having space in his car; my Spanish tandem partners, Jessica Nieto and Carlos Pascual, for the openness with which they introduced me to a new and exciting culture; my running team members from Kovošrot Praha and later from Slavia Praha, for always being very welcoming when I was back home; and everyone else, whom I may have inadvertently omitted.

Very special acknowledgements go to my family, especially my parents. It is no understatement to say that without their never-ending help and patience, I could never dream of writing this thesis.

Finally, I would like to thank Hooria Komal for her unconditional love and support, especially when I was losing faith – she knows it has happened many times, but she always managed to put me back on track.

Lausanne, December 14, 2012 L. C.

Abstract

We analyze resource allocation problems where N independent agents want to access C resources. Each resource can be only accessed by one agent at a time. In order to use the resources efficiently, the agents need to coordinate their access. We focus on decentralized coordination, i.e. coordination without central authority. We analyze coordination mechanisms for two different kind of agents: 1) cooperative, who follow the prescribed protocol entirely; and 2) non-cooperative, who attempt to maximize their own utility.

We propose a novel approach to achieve a fair and efficient resource allocation when the agents are cooperative. The agents access resources in slots. At the beginning of each slot, they observe a global coordination signal – a random integer $1 \leq k \leq K$. The agents then learn a different allocation for each value of the coordination signal. When modeled as a game, the canonical solution to the resource allocation problem is the correlated equilibrium. In a correlated equilibrium, a "smart" coordination device chooses the actions for the "stupid" agents, who then have no incentive to deviate from these recommendations. In contrast, in our solution the coordination signal is "stupid", since it is not specific to the game. The agents are "smart", because they learn their strategy independently for each signal value. We show that our learning algorithm converges to an efficient resource allocation in polynomial time. The resulting allocation becomes more fair as the number of signals K increases.

A non-cooperative, self-interested agent can exploit our cooperative allocation scheme by accessing one resource all the time, until everyone else gives up. Therefore, for the non-cooperative agents, we consider an infinitely repeated resource allocation game with discounting. This game is symmetric and all the agents are identical, so we look for its symmetric subgame-perfect equilibria. (Bhaskar (2000)) proposed a solution for 2-agent, 1-resource allocation games: Agents start by symmetrically choosing their actions randomly, and as soon as they each choose different actions, they start to follow a *convention* that prescribes their actions from then on. We extend the concept of convention to the general resource allocation problems of N agents and C resources. We show that for any convention, there exists a symmetric subgame-perfect equilibrium that implements it. We present two conventions: bourgeois, where agents stick to the first allocation; and market, where agents pay for the use of resources, and observe a global coordination signal that allows them to alternate between different allocations. We define the price of anonymity of a convention as the ratio between the maximum social payoff of any (asymmetric) strategy profile and the expected social payoff of the convention. We show that the price of anonymity of the bourgeois convention is infinite. The market convention decreases this price by reducing the conflict between the agents.

Abstract

Keywords: Resource allocation, Multi-agent Learning, Game Theory, Symmetric Equilibria, Correlated Equilibria, Efficiency, Fairness

Résumé

Nous analysons des problèmes d'allocation de ressources où N agents indépendants veulent accéder à C ressources. Chaque ressource peut être accédé par seulement un agent au même temps. Pour utiliser les ressources de manière efficace, les agents doivent coordonner leurs accès. Nous nous concentrons sur la coordination décentralisé, sans aucune autorité centrale. Nous analysons des mécanismes de coordination pour deux types d'agents : 1) coopératifs, qui suivent notre protocole sans déviations ; et 2) non coopératifs, qui essaient de maximiser leur propre utilité.

Nous proposons une nouvelle approche pour obtenir une allocation de ressources efficace et équitable quand les agents sont coopératifs. Les agents accèdent les ressources dans les créneaux uniformes. Au début de chaque créneau, ils observent un signal de coordination global – un entier aléatoire $1 \leq k \leq K$. Les agents apprennent une allocation différente pour chaque valeur de signal. Quand on modélise le problème comme un jeu, la solution canonique est l'équilibre corrélé. Dans l'équilibre corrélé, un appareil de coordination „intelligent" choisit les actions pour les agents „stupides", qui n'ont aucune motivation de dévier de ces recommandations. Au contraire, dans notre solution le signal de coordination est „stupide", car il n'est pas spécifique à notre jeu. Les agents sont quant à eux „intelligents", parce qu'ils apprennent leur stratégie pour chaque valeur de signal. Nous montrons que notre algorithme d'apprentissage converge à une allocation de ressources efficace en temps polynomial. L'allocation résultante devient plus équitable tant le nombre de signales K augmente.

Un agent non coopératif peut exploiter notre algorithme coopératif en accédant tout le temps, jusqu'à ce que les autres agents abandonnent les ressources. Pour cette raison, quand les agents sont non coopératifs, nous analysons un jeu répété à horizon infini. Ce jeu est symétrique et les agents sont identiques, nous cherchons alors son équilibre parfait de jeu partiel. Bhaskar (2000) a proposé une solution pour les jeux d'allocation avec 2 agents et 1 ressource. Les agents commencent en choisissant leurs actions de manière aléatoire, et lorsqu'ils choisissent chacun une action distincte, ils suivent une *convention* qui leurs prescrit leurs actions par la suite. Nous généralisons le concept d'une convention au problèmes génériques d'allocation de ressources avec N agents et C ressources. Nous montrons que pour chaque convention, il existe une implémentation symétrique qui est un équilibre parfait d'un jeu partiel. Nous présentons deux conventions : la bourgeoise, où les agents gardent la première allocation efficace qu'ils ont obtenus ; et la convention de marché, où les agents paient pour l'utilisation de ressources, et où ils observent un signal de coordination qui leur permet d'alterner parmi plusieurs allocations. Nous définissons le prix d'anonymat d'une

convention comme la proportion entre le profit social maximal d'une stratégie asymétrique, et le profit social de la convention. Nous prouvons que le prix d'anonymat de la convention bourgeoise est infini. La convention de marché réduit ce prix en diminuant le conflit entre les agents.

Contents

List of Figures

List of Tables

List of Symbols

A	Access action – a player accesses a resource chosen uniformly at random	9
A_c	Action to access resource c	17
C	Number of resources in a resource allocation game	17
$E(\sigma)$	Expected social payoff of strategy profile σ	22
E_A	Expected payoff function for taking action A (access a resource)	57
E_Y	Expected payoff function for taking action Y (yield)	57
E_i	Expected future discount payoff of player i	12
$G_{N,C}$	Resource allocation game of N agents and C resources	17
H^A	Hitting time of a subset $A \subset I$ of the states of a Markov chain	15
I	Set of states of a Markov chain	14
$J(\mathbb{X})$	Jain index of the resource allocation $\mathbb{X}$	32
K	Number of coordinations signals	26
N	Number of players in a game	7
R_G	Price of anonymity of a symmetric game G	22
$R_G(\sigma)$	Price of anonymity of a strategy vector σ in a symmetric game G	22
X_t	State of a Markov chain at time t	14
Y	Yield action – a player does not access any resource in the resource allocation game	9
$\Delta(\mathscr{A}_i)$	Space of probability distributions over action space $\mathscr{A}_i$	8
Ω	Probability space	14
χ_i	Strategy of player i in a repeated game $\mathscr{G}$	12
δ	Discount factor	11

1 Introduction

In many situations, there are several users who want to use a resource that can be successfully used only by one user at a time. In a simple wireless network model, only one device may transmit on a given channel. If multiple devices attempt to transmit at the same time, their transmissions interfere with each other and fail. Similarly, one parking lot may only be used by one vehicle at a time, lest there be a traffic accident. And last but not least, when bidding for one item in several simultaneous auctions, a bidder prefers the auction with less participants, because this will usually lead to a lower price.

In all of these situations, all the users of a resource (we will call them the *agents*) prefer the same outcome: to be the only one who accesses the resource. However, if several agents access a given resource at the same time, they collide, and they are worse off than had they not accessed the resource in the first place. Therefore, it is useful to provide a coordination mechanism in order to achieve an efficient use of the resources. Such coordination can have two principal forms:

Centralized where there is a central authority that tells users when to access which resource, and

Decentralized where there is no such authority and all the agents have to adopt a common protocol that helps them access the right resource at the right time.

The resources that we consider are homogeneous, i.e. they are all identical. The agents have also identical preferences – they are indifferent which resource they get.

Centralized approaches (such as *Time-division multiple access, (TDMA)*), offer high efficiency and fair resource allocation. This is because the central authority can make sure that only one user accesses a given resource at a time. The central authority can also alternate between users so that they can all access a resource equally often. However, the problem of the centralized solution is that the central authority might not always be available. Also, the central authority is a single point of failure of the resource allocation mechanism – if it stops functioning, the

resource allocation may break down. Therefore, in this work, we will focus on decentralized resource allocation mechanisms.

There are two main settings for decentralized resource allocation:

Cooperative where all the agents adopt a given protocol and follow it completely.

Non-cooperative where every agent tries to maximize its own utility, without regard for the others.

In the cooperative setting, all the agents want to achieve the most efficient resource allocation overall – they don't care about their own utility *per se*. They trust the protocol designer to design the allocation algorithm to be as efficient as possible. An example of a simple protocol for cooperative wireless channel allocation is *ALOHA* (Abramson (1970)). There, multiple devices attempt to transmit data over one shared channel. Access to the channel is slotted, that is, the transmissions can only start at the beginning of a given time interval. If multiple devices transmit in the same slot, the transmissions collide and fail. According to the ALOHA protocol, after a collision, the devices wait for a random period of time before they retransmit. Because the devices decide randomly when to transmit, it may happen that there will be a collision again. If there are N devices who try to transmit over a shared channel, the highest use of the channel is achieved when each device transmits with probability $\frac{1}{N}$. Asymptotically, the resulting throughput (i.e. percentage of time the channel is used for communication) converges to $\frac{1}{e} \approx 37\%$ (Alam and Hossain (1997)).

In the non-cooperative setting, each agent tries to maximize her own utility. Imagine a problem of allocating channels for multiple wireless networks in the same apartment building. Each apartment has its own wireless network that is managed independently. Standard wireless protocols (such as 802.11b/g) only specify a handful of non-overlapping wireless channels (usually 3) that can be used independently. Within a single network, all devices use the same channel and the same access point. This access point can allocate the wireless channel centrally, using a mechanism such as TDMA. However, when there are multiple independent networks close to each other, they can interfere if too many networks use the same channel. Since there is no such central authority to decide which network is going to use which channel, the networks have to choose the channels themselves. Naturally, each network attempts to choose the channel with the least interference. What happens if all networks attempt to do the same?

Non-cooperative settings are usually analyzed using game theory. Game theory is a branch of economics that considers strategic interactions between rational, self-interested agents. It analyses the behaviour of such rational agents in these interactions, and it attempts to predict the interaction outcome. It should be *rational* for the agents to adopt such outcome. Usually, we say that an outcome is rational if it forms some kind of equilibrium. In an equilibrium, no agent can improve her utility by acting differently. The most essential equilibrium is the

so-called *Nash equilibrium* (Nash (1951)), where no agent can improve her utility by taking a different action, provided that the other agents keep playing the same actions.

There are three desirable properties of a good resource allocation protocol:

Efficiency All the resources are used in a useful way, all the time;

Fairness All the agents can use some resource equally often;

Rationality No agent can increase her utility by deviating from the prescribed protocol; in the terminology of game theory, the protocol is an equilibrium of the resource allocation game.

Efficiency and fairness is a concern in both cooperative and non-cooperative settings. However, as we stated above, the agents in the cooperative setting are not rational, since they are not interested in maximizing their own utility. Therefore, rationality is only a concern in the non-cooperative setting.

Contributions on Cooperative Resource Allocation

Existing mechanism for decentralized resource allocation in the cooperative setting can be classified into two categories. First, there are algorithms where the agents rely on simple feedback on whether their own access was successful or not. The agents usually start by accessing the resource randomly, and then try to learn from the collisions with other agents (such as in the ALOHA protocol described above). While these algorithms are simple to implement and achieve a fair resource allocation, they are not efficient. The second category of algorithms are algorithms based on explicit coordination using some form of message exchange. Such algorithms are based for example on *Distributed Constraint Optimization (DCOP)* (Cheng et al. (2009); Modi et al. (2002)) or *Generalized Partial Global Planning* (Decker and Li (1998)). After an initial message exchange, such algorithms can achieve an efficient and fair allocation of the resources. However, the messages create an overhead that can eliminate their benefits.

Ideally, we would like to achieve the (nearly) efficient use of resources, with no explicit communication between the agents. In this thesis, we propose to use a simple coordination signal that all the agents can observe and that fluctuates ergodically. This signal could be a common clock, radio broadcasting on a specified frequency, the decimal part of a price of a certain stock at a given time, etc. Depending on this “stupid” signal, “smart” agents can then learn to take a different action for each of its value.

We present a learning algorithm that achieves an efficient allocation in time polynomial in the number of agents and resources for each value of the coordination signal. The agents only receive binary feedback – when they access a resource, they learn whether their access was

successful or whether they collided. When they don't access any resource, they can choose a resource to observe. They then learn if this resource was accessed by some agents or not. The agents cannot communicate with each other. As the number of coordination signals increases, the overall allocation becomes more fair. We experimentally evaluate how sensitive the algorithm is to a player population that is dynamic, i.e. when players enter and leave the system. We also evaluate the algorithm's resistance to noise, be it in the feedback players receive or in the coordination signal they observe.

Contributions on Non-cooperative Resource Allocation

The second issue we analyze in this thesis is the problem of resource allocation in a non-cooperative setting. The algorithms for the cooperative setting assumed that the agents follow the prescribed protocol entirely. However, a self-interested agent may gain an advantage by deviating from the protocol. For example, in the ALOHA protocol, a self-interested agent may transmit over a wireless channel all the time, until everyone else learns not to transmit. The deviating agent then claims the wireless channel for herself only. The other agents would have been better off not transmitting at all.

We want to limit ourselves to protocols that are *rational* for self-interested agents to adopt. In the language of game theory, such a protocol is an *equilibrium* of the resource allocation problem. Ideally, such an equilibrium should be as efficient as possible. The efficiency is measured by how often is the resource used by one agent only. The problem is that many efficient equilibria are asymmetric – they assign the resources only to a fixed subset of agents. There are two problems with asymmetric equilibria. First, they are not fair – the other agents are never allowed to access a resource. Second, they pose a problem for coordination. At the beginning, when none of the resources have been assigned yet, all the agents may believe they have a right to use them. Consequently, the agents will collide.

Consider the following example: Millions of wireless sensors are produced all by the same pipeline. We take two of them randomly, and put them in a room. There is only one frequency on which the sensors can transmit their measurements. How can each sensor know when to transmit and when to stay quiet? The factory could program half of the sensors to transmit in odd slots, and the other half to transmit in the even slots. Nevertheless, it would be just as likely to have an odd-even pair of sensors, as it would be to have a pair where the sensors transmit at the same time.

Therefore, we limit ourselves only to symmetric equilibria of the resource allocation problem. Previous work in game theory (Bhaskar (2000); Kuzmics et al. (2010)) considered problems simpler than the general resource allocation with N agents and C resources. They assumed that the agents allocate resource in time slots of fixed length, and they proposed a simple symmetric protocol: The agents start by accessing some resource randomly with a certain probability, and as soon as they "stumble upon" an efficient allocation, they follow a *convention* that prescribes the allocation in the next slot based on the allocation in the past slots.

Limiting ourselves to only symmetric equilibria of the resource allocation game comes at a cost: the most efficient equilibria may be (and often are) asymmetric. To measure the efficiency of a given symmetric equilibrium, we define the *price of anonymity*. It is the ratio between the highest expected payoff of any asymmetric protocol, and the social payoff of the given symmetric equilibrium protocol in question.

We formally define the convention for the N-agent, C-resource allocation problems, and we show that any *equilibrium* convention can be implemented as an equilibrium of the resource allocation game. We then present two conventions: the *bourgeois* and the *market* convention. In the bourgeois convention, the agents who have successfully accessed some resource keep accessing the same resource forever. We show that when the number of resources is small relative to the number of agents, the expected payoff the agents get is zero. The market convention improves the coordination between the agents by reducing the conflict between them. This convention increases the resource supply by using a global coordination signal, just like in our cooperative solution. At the same time, it decreases the demand for resources by charging the agents a fixed price for each successful access. We show that the price of anonymity of the market convention is finite.

Outline of this Thesis

In Chapter 2, we introduce some basic background concepts that we will use throughout the thesis. We define the basic notions of the game theory and show how we can model the resource allocation problem in the framework of the game theory. We then review the theory of Markov chains, that we will use when proving the convergence of our algorithms. We review the previous literature on symmetric equilibria of symmetric games, and we extend the concept of convention to the general resource allocation game. We define the price of anonymity as a measure of how efficient symmetric equilibria are compared to asymmetric ones.

In Chapter 3, we describe the distributed protocol for resource allocation, based on the global coordination signal. We formally prove its convergence, and we show that it reaches an (almost) fair allocation. We present the results of our simulations using this protocol.

Since it is not rational for self-interested agents to adopt the protocol from Chapter 3, we discuss some ways to construct rational protocols that are also efficient and fair in Chapter 4. We extend the concept of convention to *augmented* convention, that allows the agents use a global coordination signal. We show that in the resource allocation game, for any equilibrium convention there exists an equilibrium *implementation* such that when the agents follow the implementation and then the convention, they play a symmetric subgame-perfect equilibrium of the resource allocation game. We present the *bourgeois* and the *market* convention, and analyze their efficiency and convergence.

Finally, in Chapter 5 we conclude and we present possible future applications of this work.

2 Preliminaries

This chapter presents the general background knowledge on which our work is based on. In Section 2.1, we present the basic notions of the game theory. In Section 2.2 we present the theory of Markov chains that we will use later to analyze the performance of our algorithms theoretically.

Section 2.3 presents the concept of a *convention* that allows us to reach asymmetric outcomes using a symmetric strategy. Finally, in Section 2.4, we present the *price of anonymity*. It is a measure of how efficient a symmetric strategy is compared to the most efficient asymmetric one.

2.1 Game theory

Game theory is the study of interactions among independent, self-interested agents. An agent who participates in a game is called a *player*. Each player has a utility function associated with each state of the world. Self-interested players take actions so as to achieve a state of the world that maximizes their utility. Game theory studies and attempts to predict the behaviour, as well as the final outcome of such interactions.

The basic way to represent a strategic interaction (*game*) is using the so-called *normal form* (the following definitions are cited from Leyton-Brown and Shoham (2008)).

Definition 1. A finite, N-person *normal-form game* is a tuple $(\mathbf{N}, \mathscr{A}, u)$, where

- $\mathbf{N}$ is a set of N players ;
- $\mathscr{A} = \mathscr{A}_1 \times \mathscr{A}_2 \times \ldots \times \mathscr{A}_N$, where $\mathscr{A}_i$ is a set of actions available to player i. Each vector $\mathbf{a} = (a_1, a_2, \ldots, a_N) \in \mathscr{A}$ is called an *action profile* ;
- $u = (u_1, u_2, \ldots, u_N)$, where $u_i : \mathscr{A} \to \mathbb{R}$ is a *utility function* for player i that assigns each action vector a certain utility (payoff).

	C	D
C	3, 3	1, 4
D	4, 1	2, 2

Figure 2.1: Example of a game in the normal form: The Prisoners' dilemma. Both players have two actions: C, "cooperate", and D, "deviate".

In practice, we can represent the normal form of a game as an N-dimensional matrix, where each cell contains the utility vector for all the players for a given action vector.

Figure 2.1 shows an example of the normal form representation of (arguably) the most famous game – the Prisoners' dilemma. In this game, there are two players, who have each two actions: to cooperate (C), or defect (D). Both players prefer for both to cooperate rather than both defect. However, the most most beneficial for each player is to defect on her own and let the other player cooperate. In that case, the defecting player receives a payoff of 4, while the cooperating player receives a payoff of only 1.

When playing a game, players have to select their *strategy*. A *pure* strategy σ_i for player i selects only one action $a_i \in \mathscr{A}_i$. A vector of pure strategies for each player $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ is called a *pure strategy profile*. A *mixed* strategy selects a probability distribution over the entire action space, i.e. $\sigma_i \in \Delta(\mathscr{A}_i)$. A *mixed strategy profile* is a vector of mixed strategies for each player. For a mixed strategy σ_i, we define its support $supp(\sigma_i)$ as

$$supp(\sigma_i) = \{a_i \in \mathscr{A}_i : \sigma_i(a_i) > 0\}.$$

That is, the support of a mixed strategy is a set of actions that the strategy plays with non-zero probability.

The question is, for a given game specified using its normal form, how should the players choose their strategy? What are the desirable outcomes? One way to choose is using the concept of *Pareto efficiency*:

Definition 2. Strategy profile σ *Pareto dominates* strategy profile σ' if for all players $i \in \mathbf{N}$, $u_i(\sigma) \geq u_i(\sigma')$, and there exists some $j \in \mathbf{N}$ such that $u_j(\sigma) > u_j(\sigma')$.

Definition 3. Strategy profile σ is *strictly Pareto optimal* (or *efficient*), if there does not exist any other strategy profile $\sigma' \neq \sigma$ such that σ' Pareto dominates σ.

Another way to choose the outcome to play is to compare the *social payoff* of a given outcome:

Definition 4. For an action vector $(a_1, a_2, \ldots, a_N)$, we define its *social payoff* as the sum of utilities of all the players, $\sum_{i=1}^{N} u_i(a_1, a_2, \ldots, a_N)$.

	Y	A
Y	0, 0	0, 1
A	1, 0	-2, -2

Figure 2.2: The game of *Chicken*. The players have two actions: Y, *"yield"*, and A, *"access"*.

When players know the strategies of the others, they can also choose their action quite easily: just pick the strategy that maximizes the payoff *given* what everyone else is playing:

Definition 5. We say that a mixed strategy σ_i^* of player i is a *best response* to the strategy profile of the opponents σ_{-i} if for any strategy σ_i',

$$u_i(\sigma_i^*, \sigma_{-i}) \geq u_i(\sigma_i', \sigma_{-i})$$

As we mentioned earlier, one of the basic goals of game theory is to predict an outcome of a strategic interaction. Such outcome should be stable – therefore, it is usually called an *equilibrium*. One requirement for an outcome to be an equilibrium is that none of the players has an incentive to change their strategy, i.e. all players play their best-response to the strategies of the others. This defines perhaps the most important equilibrium concept, the Nash equilibrium:

Definition 6. A strategy profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ is a *Nash equilibrium* (NE) if for every player i, her strategy σ_{-i} is a best response to the strategies of the others σ_{-i}.

The importance of the Nash equilibrium stems from the fact that it is always guaranteed to exist:

Theorem 1. *(Nash (1951)) Every game with a finite number of players and action profiles has at least one Nash equilibrium.*

If we consider the Prisoners' dilemma game from Figure 2.1, it has one Nash equilibrium: Both players play action D (*"deviate"*). This is because when player 2 plays D, the player 1 gets payoff of 1 when playing C, and payoff of 2 when playing D. Playing D is therefore the best response. We say that this equilibrium is a *pure strategy* Nash equilibrium (*PSNE*).

A game may have more than one Nash equilibrium. Consider the game of *Chicken* from Figure 2.2. It simulates the situation where two players try to access a resource (such as a parking lot, wireless channel, etc.). Each player has two actions: to *access* (A) the resource, and to *yield* (Y).

The game of Chicken has two pure-strategy Nash equilibria: (A, Y) and (Y, A). It has also one *mixed strategy* Nash equilibrium (*MSNE*), where each player chooses to play the action A with probability $\frac{1}{3}$, and action Y also with probability $\frac{2}{3}$.

For 2-player games, we can calculate the mixed strategy Nash equilibrium as follows:

- Pick the supports of the mixed strategies $\mathscr{B}_1 \subset \mathscr{A}_1$ and $\mathscr{B}_2 \subset \mathscr{A}_2$.
- The equilibrium strategy σ_1 for player 1 is a probability distribution over $\mathscr{B}_1$ such that the expected payoff for player 2 when playing any action from the set $\mathscr{B}_2$ is the same. This way, player 2 will be indifferent between playing actions in $\mathscr{B}_2$.
- Then, find the equilibrium σ_2 for player 2 the same way.
- If both probability distributions exist, the mixed strategy profile (σ_1, σ_2) is a mixed strategy Nash equilibrium.

In our example of the Chicken game, when player 2 adopts the mixed strategy of playing A with probability $\frac{1}{3}$, the expected payoffs to player 1 are the following:

$$E_1(A|\sigma_2) := \frac{1}{3} \cdot (-2) + \frac{2}{3} \cdot 1 = 0, \tag{2.1}$$

$$E_2(Y|\sigma_2) := \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 0 = 0. \tag{2.2}$$

As essential as Nash equilibria are, they have several disadvantages. First, Finding a Nash equilibrium of any N-player game (specified in its normal form) is *PPAD-complete* (Chen and Deng (2006)); that means that unless $P = NP$, there is no algorithm to find a Nash equilibrium with runtime polynomial in the size of the game in the worst case. Second, as we saw in the example of the Chicken game, there might be several Nash equilibria. Which equilibrium should the players adopt? Third, the most efficient Nash equilibrium may not be the most fair one. In the Chicken game, the two PSNEs have a social payoff of 1. However, they are not fair, because one of the players gets much worse payoff than the other. On the other hand, the mixed strategy Nash equilibrium is fair, but has only a social payoff of 0.

Aumann (1974) proposed the *correlated equilibrium* (*CE*) that fixes these issues. Intuitively, a correlated equilibrium is a probability distribution over the joint strategy profiles of the game. Before the game, a correlation device samples this distribution and recommends to each player an action to play. The probability distribution is a CE if the players don't have an incentive to deviate from the recommended action.

The formal definition of the CE is as follows:

Definition 7. Given an N-player game $(\mathbf{N}, \mathscr{A}, u)$, a *correlated equilibrium* is a tuple (v, π, μ), where v is a tuple of random variables $v = (v_1, v_2, \ldots, v_N)$ with domains $D = (D_1, D_2, \ldots, D_N)$,

π is a joint probability distribution over v, $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$ is a vector of mappings $\mu_i : D_i \mapsto \mathscr{A}_i$, and for each player i and every mapping $\mu'_i : D_i \mapsto \mathscr{A}_i$ it is the case that

$$\sum_{d \in D} \pi(d) u_i \left(\mu_1(d_1), \mu_2(d_2), \ldots, \mu_N(d_N)\right) \geq \sum_{d \in D} \pi(d) u_i \left(\mu'_1(d_1), \mu'_2(d_2), \ldots, \mu'_N(d_N)\right).$$

Correlated equilibria have several nice properties [1] For any game, the set of correlated equilibria forms a superset of the set of Nash equilibria:

Theorem 2. *(Leyton-Brown and Shoham (2008)) For every Nash equilibrium σ^*, there exists a corresponding correlated equilibrium (v, π, μ).*

The set of correlated equilibria is convex (as shown by Fudenberg and Tirole (1991)), so the set of correlated equilibria is at least as large as the convex hull of the Nash equilibria. Correlated equilibria are also easier to find than Nash equilibria: We can find a correlated equilibrium of a given game in polynomial time using linear programming (Papadimitriou and Roughgarden (2008)).

In the Chicken game, there is a correlated equilibrium that is efficient and fair: the global correlation device selects one of the two pure strategy Nash equilibria with probability 0.5 and then tells each player whether to access or yield. The expected social payoff is 1, and the expected payoff to any of the player is 0.5.

2.1.1 Repeated games

In a *repeated game*, the same players play a given game (for example specified by its normal form) repeatedly. We call the game that is being played the *stage game*.

Definition 8. Given an infinite sequence of payoffs $r_i^{(1)}, r_i^{(2)}, \ldots$ for player i and a discount factor δ , $0 < \delta < 1$, the *future discounted reward* of player i is

$$\sum_{j=1}^{\infty} \delta^j r_i^{(j)}$$

Definition 9. Let $G = (N, A, u)$ be a normal form game. And *infinitely repeated* version $\mathscr{G}$ of the game G *with discounting* is a game where the players play the normal form game G for an infinite number of rounds. The players discount future payoff with a discount factor δ.

The are two interpretations of the discount factor δ: Either it expresses the fact that each player cares more about the current round of the game than about the future rounds. Or, it is the probability that the game will continue after each round.

[1] Roger Myerson, a Nobel-prize winning economist, has been quoted saying that *"If there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium."* (Myerson (1997))

Definition 10. Let $\mathcal{G}$ be an infinitely repeated game with discounting. We define the *history* h_t of the play in round $t \geq 0$ as

$$h_t := \left((a_1^0, a_2^0, \ldots, a_N^0), \ldots, (a_1^{t-1}, a_2^{t-1}, \ldots, a_N^{t-1})\right)$$

where a_i^t is the action taken by player i in round t.

We define $\mathcal{H}$ the space of a all possible histories of the game $\mathcal{G}$.

Definition 11. A *strategy in the infinitely repeated game* of an player i is a function from the history to a probability distribution over the action space,

$$\chi_i : h_t \mapsto \Delta(\mathcal{A}_i)$$

For a given strategy profile of the symmetric game $\chi = (\chi_1, \chi_2, \ldots, \chi_N)$, the expected utility function to player i is

$$u_i(\chi) = \sum_{t=0}^{\infty} \delta^t u_i((\chi_1(h_t), \ldots, \chi_N(h_t))). \tag{2.3}$$

We can define the Nash equilibrium of the repeated game in the same way as for the stage game (we can treat the repeated game as if it was just a normal form game where players commit to their strategy for the entire game up front).

Definition 12. A strategy profile $\chi = (\chi_1, \chi_2, \ldots, \chi_N)$ is a *Nash equilibrium of the infinitely repeated game* if for each player i,

$$u_i(\chi_i, \chi_{-i}) \geq u_i(\chi_i', \chi_{-i}) \tag{2.4}$$

for any alternative strategy of the repeated game χ_i'.

For the repeated games, there exists a stronger notion of equilibria, which is a refinement of the standard Nash equilibrium definition.

Definition 13. Let $\mathcal{G}$ be an infinitely repeated game with a discount factor $0 < \delta < 1$. A strategy vector $\chi = (\chi_1, \chi_2, \ldots, \chi_N)$ is a *subgame-perfect equilibrium* of the game $\mathcal{G}$ if for each player i,

$$E_i((\chi_i, \chi_{-i}), h_t) \geq E_i((\chi_i', \chi_{-i}), h_t)$$

for any strategy χ_i' and history h_t. Here $E_i((\chi_i, \chi_{-i}), h_t)$ is the future discounted payoff of player i when she adopts strategy χ_i and the other players adopt a strategy vector χ_{-i}.

In the subgame-perfect equilibrium, players play a best-response strategy given any history of the play, including the histories which cannot occur if they follow the equilibrium strategy from

the beginning. The notion of subgame-perfect equilibria eliminates this way "non-credible threats", or equilibria in which a player threatens someone else with a strategy which the player might be prefer to avoid if it was supposed to be executed.

For an infinitely repeated game with discounting, the set of Nash equilibria can be characterized using the so-called *folk theorem*. While their name indicates that they have been known and used well before they were first published, we will follow the version described by Fudenberg and Maskin (1986). Informally, the folk theorem states that in the infinitely repeated game, for every feasible and individually rational payoff vector of the stage game, there exists a Nash equilibrium of the repeated game where the average payoffs per round correspond to the stage game payoff vector.

A payoff vector is individually rational if it Pareto-dominates the minimax payoff of the stage game. For player i, the minimax payoff is

$$v_i^* := \min_{\sigma_{-i}} \max_{\sigma_i} u_i(\sigma_i, \sigma_{-i}). \tag{2.5}$$

To simplify the notation, Fudenberg and Maskin normalize the payoffs so that $(v_1^*, v_2^*, \ldots, v_N^*) = (0, 0, \ldots, 0)$. Let

$$U := \{(v_1, \ldots, v_N) | \exists (a_1, \ldots, a_N) \in \mathscr{A}_\infty \times \ldots \times \mathscr{A}_\mathscr{N} \text{ s.t. } u(a_1, \ldots, a_N) = (v_1, \ldots, v_N)\},$$

$$V := \text{convex hull of } U,$$

$$V^* := \{(v_1, \ldots, v_N) \in V | v_i > 0 \text{ for all } i\}.$$

The set V is the set of feasible payoffs in the stage games (that is, payoffs which can be achieved by playing some mixed or correlated strategy). The set V^* is the subset of feasible payoffs which are also individually rational.

Theorem 3. *(Fudenberg and Maskin (1986)) For any $(v_1, \ldots, v_N) \in V^*$, if the discount factor δ is close enough to 1, there exists a Nash equilibrium of the infinitely repeated game with discounting where, for all i, the average payoff to player i is v_i.*

The idea of the proof is as follows: The agents cycle through a prescribed sequence of game outcomes so that they achieve the desired payoffs. If one player deviates, the others punish him by playing the minimax strategy forever after.

2.2 Markov chains

Many of the decision strategies that we propose and analyze in this work can be described as randomized algorithms. In a randomized algorithm, some of its steps depend on the value of a random variable. One useful technique to analyze randomized algorithms is to describe its

execution as a *Markov chain*.

A Markov chain is a random process with the *Markov property*. A random process is a collection of random variables; usually it describes the evolution of some random value over time. A process has a Markov property if its state (or value) in the next time step depends exclusively on its value in the previous step, and not on the values further in the past. We can say that the process is *memoryless*. If we imagine the execution of a randomized algorithm as a finite-state automaton with non-deterministic steps, it is easy to see how its execution maps to a Markov chain.

The formal definition of a Markov chain is as follows:

Definition 14. (Norris (1998)) Let I be a countable set. Each $i \in I$ is called a *state* and I is called the *state space*. We say that $\lambda = (\lambda_i : i \in I)$ is a *measure* on I if $0 \leq \lambda_i < \infty$ for all $i \in I$. If in addition the *total mass* $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a *distribution*. We work throughout with a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Recall that a *random variable* X with values in I is a function $X : \Omega \to I$. Suppose we set

$$\lambda_i = \Pr(X = i) = \Pr(\{\omega \in \Omega : X(\omega) = i\}).$$

Then λ defines a distribution, the distribution of X. We think of X as modelling a random state that takes value i with probability λ_i.

We say that a matrix $P = (p_{ij} : i, j \in I)$ is *stochastic* if every row $(p_{ij} : j \in I)$ is a distribution.

We say that $(X_t)_{t \geq 0}$ is a *Markov chain* with *initial distribution* λ and a *transition matrix* P if

1. X_0 has distribution λ;
2. for $t \geq 0$, conditional on $X_t = i$, X_{t+1} has distribution $(p_{ij} : j \in I)$ and is independent of $X_0, X_1, \ldots, X_{t-1}$.

More explicitly, the conditions state that, for $t \geq 0$ and $i_0, \ldots, i_{t+1} \in I$,

1. $\Pr(X_0 = i_0) = \lambda_{i_0}$;
2. $\Pr(X_{t+1} = i_{t+1} | X_0 = i_0, \ldots, X_t = i_t) = p_{i_t i_{t+1}}$.

Theorem 4. *Let A be a set of states. The vector of hitting probabilities $h^A = (h_i^A : i \in \{0, 1, \ldots, N\})$ is the minimal non-negative solution to the system of linear equations*

$$h_i^A = \begin{cases} 1 & \text{for } i \in A \\ \sum_{j \in \{0,1,\ldots,N\}} p_{ij} h_j^A & \text{for } i \notin A \end{cases}$$

One property of randomized algorithms that we are particularly interested in is its convergence. If we have a set of states A where the algorithm has converged, we can define the time it takes

to reach any state in the set A from any other state of the corresponding Markov chain as the *hitting time*:

Definition 15. (Norris (1998)) Let $(X_t)_{t\geq 0}$ be a Markov chain with state space I. The *hitting time* of a subset $A \subset I$ is a random variable $H^A : \Omega \to \{0,1,\ldots\} \cup \{\infty\}$ given by

$$H^A(\omega) = \inf\{t \geq 0 : X_t(\omega) \in A\}$$

Specifically, we are interested in the *expected* hitting time of a set of states A, given that the Markov chain starts in an initial state $X_0 = i$. We will denote this quantity

$$k_i^A = \mathbb{E}_i(H^A).$$

In general, the expected hitting time of a set of states A can be found by solving a system of linear equations.

Theorem 5. *The vector of expected hitting times $k^A = E(H^A) = (k_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} k_i^A = 0 & \text{for } i \in A \\ k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A & \text{for } i \notin A \end{cases} \tag{2.6}$$

Convergence to an absorbing state may not be guaranteed. To calculate the probability of reaching an absorbing state, we can use the following theorem (Norris (1998)):

Theorem 6. *Let A be a set of states. The vector of hitting probabilities $h^A = (h_i^A : i \in \{0,1,\ldots,N\})$ is the minimal non-negative solution to the system of linear equations*

$$h_i^A = \begin{cases} 1 & \text{for } i \in A \\ \sum_{j \in \{0,1,\ldots,N\}} p_{ij} h_j^A & \text{for } i \notin A \end{cases}$$

Solving the systems of linear equations in Theorems 5 and 6 analytically might be difficult for many Markov chains though. Fortunately, when the Markov chain has only one absorbing state $i = 0$, and it can only move from state i to j if $i \geq j$, we can use the following theorem to derive an upper bound on the expected hitting time (proved by Rego (1992)):

Theorem 7. *Let $A = \{0\}$. If*

$$\forall i \geq 1 : E(X_{t+1} | X_t = i) < \frac{i}{\beta}$$

for some $\beta > 1$*, then*

$$k_i^A < \left\lceil \log_\beta i \right\rceil + \frac{\beta}{\beta - 1}$$

2.3 Conventions

As we have shown for the example of the game of Chicken from Figure 2.2, there exist symmetric games that have nevertheless only asymmetric efficient equilibria. If we allow for a central coordination device, the agents can play a symmetric and efficient correlated equilibrium that selects randomly from the set of efficient Nash equilibria. Without such a device, in the stage game, there is no way to reach a symmetric efficient outcome in an equilibrium.

However, if the agents play the game repeatedly, they can use the history of the play to condition their strategy. If two agents have different histories, they can take different actions in the future. In the first round of the game though, the history is empty for everyone. Therefore, a symmetric strategy for the players has to randomize in order to ever reach a point when the histories of the agents are distinct.

2.3.1 Previous work

Bhaskar (2000) considered the problem of playing asymmetric outcomes of the stage game using a symmetric strategy of the repeated game. His work considers games with 2 players and 2 actions, such as the game of Chicken. The idea is that the two players start by playing randomly, using the same probability distribution over actions. They randomize until they reach a round t where they happen to play some pure-strategy Nash equilibrium (that is, they take a different action each). We call this round the *asynchrony* round. Then, the agents start following a so-called *convention*. A convention maps the asymmetric pure-strategy Nash equilibrium to a (potentially asymmetric) strategy vector that the agents then adopt.

Bhaskar gives two examples of a convention for the 2-player, 2-actions game:

Bourgeois After an asynchrony, the agents keep using the action they played in the last round;

Egalitarian Agents play the action of their opponent from the last round, after asynchrony.

In the game of Chicken, in the asynchrony round, one agent chooses action A and the other one chooses Y. We will call the agent who chose A in the asynchrony round the *winner*. The other agent is the *loser*. The bourgeois convention guarantees that the agents will keep playing this NE forever after. This way, the winner will be forever guaranteed a higher payoff than the loser. In the egalitarian convention, the players alternate between the two pure-strategy Nash equilibria. That way the payoffs of the winner and a loser will be closer.

In the infinitely repeated game with discounting, the social payoff will depend on two things:

the discount factor δ, and the probability of a collision, that is the probability that the players play both action A. When there is a big difference between the winner and loser payoff, the losers will "fight back" harder, so they will play their most preferred action A with higher probability. This will increase the probability of a collision. In the egalitarian convention, the payoffs to the loser are closer to the winner. Therefore, the agents will collide less often, and they will also reach the asynchrony faster.

As another example of a convention, Kuzmics et al. (2010) analyze the *Nash demand game*. The Nash demand game is a game of N players who choose between N actions labeled $1, \ldots, N$. If all the players choose a distinct action, each player receives a payoff equal to the label of her chosen action. If there are any two players who chose the same action, every player (including those who chose an action alone) receives zero payoff. In a pure-strategy Nash equilibrium, all the players choose a different action. Naturally, each player prefers the equilibrium where she is the one who chose action N.

In the Nash demand game, we can also define bourgeois and egalitarian conventions. Kuzmics et al. define three notions of *payoff symmetry*:

Ex-ante All agents have the same expected payoffs before the game starts.

Ex-post All agents have the same expected payoffs when asynchrony occurs (regardless of who was the winner).

Strong ex-post All agents have the same payoff along any realization of the play.

The bourgeois convention is only ex-ante payoff symmetric, since once asynchrony occurs, the winner gets a higher payoff than the loser. The egalitarian convention is strong ex-post payoff symmetric. In fact, Kuzmics et al. show that in the Nash demand game, if a convention is socially efficient, it must be strong ex-post payoff symmetric. The intuition is that in order to maximize social efficiency, we want to reach asynchrony as fast as possible. This is only possible if agents choose their actions uniformly at random. They will only do that if they are indifferent between which action they choose at the moment asynchrony occurs.

2.3.2 Conventions for resource allocation games

We will formally define the resource allocation game that we analyze in this thesis as follows:

Definition 16. A *resource allocation game* $G_{N,C}$ is a game of N agents. Each agent i can access one of C resources. The agent chooses its action a_i from $\mathscr{A}_i = \{Y, A_1, A_2, \ldots, A_C\}$, where action $a_i = Y$ means to yield, and action $a_i = A_c$ means to access resource c. Because all resources are identical, we can define a special meta-action $a_i = A$. To take action A means to choose to access, and then to choose the resource uniformly at random from the set of available resources.

The payoff function for agent i is defined as follows:

$$u_i(a_1,\ldots,a_i,\ldots,a_N) := 0 \text{ if } a_i = Y \tag{2.7}$$

$$u_i(a_1,\ldots,a_i,\ldots,a_N) := \begin{cases} 1 & \text{if } a_i \neq Y, \\ & \forall j \neq i, a_j \neq a_i \\ -\gamma < 0 & \text{otherwise} \end{cases} \tag{2.8}$$

This game has a set of asymmetric pure strategy NEs where C agents each access a resource c_i and $N-C$ agents do not. There is also a symmetric mixed strategy NE where each agent decides to access *some* resource with probability

$$\Pr(a_i > 0) := \min\left\{C \cdot \left(1 - \sqrt[N-1]{\frac{|\gamma|}{1+|\gamma|}}\right), 1\right\} \tag{2.9}$$

and then chooses the resource to access uniformly at random. Note that for high enough values of C, all agents will choose to access *some* resource.

For a small number of resources C, the symmetric mixed strategy NE has an expected payoff of 0. Therefore, we will look for the symmetric equilibria of the repeated game. We will follow the same pattern as Bhaskar and Kuzmics et al. – the agents first randomize, and then adopt an asymmetric convention.

Both examples of a convention that we described above have a one common feature. All the agents choose their actions according to the same probability distribution in every round until they stumble upon a pure-strategy NE (i.e., they reach the asynchrony). The problem with this approach is that it may take a long time before the agents reach the asynchrony. Take as an example the Nash demand game of N agents of Kuzmics et al.. Any pure-strategy NE corresponds to a permutation of the set $\{1,2,\ldots,N\}$. If each agent selects action $\mathscr{A}_i$ with uniform probability $\frac{1}{N}$, the probability of playing a pure-strategy NE in any given round is $\frac{N!}{N^N}$. This means that the expected number of rounds before a pure-strategy NE is reached is $\frac{N^N}{N!} \approx O(N^N)$.

Fortunately, in the resource allocation game we can do better. To speed up convergence, we propose to learn to play the pure-strategy NE step-by-step. The agents who already happen to play their NE action alone will keep playing it until the pure-strategy NE is reached. In the resource allocation game described above, this means that the agents who successfully access some resource (we call them *winners*) will keep accessing this resource until the asynchrony round (we say that they *claim* a resource). The other agents (called the *losers*) have no incentive to access these occupied resources, since that would only lead to a collision and a negative payoff.

We will now formally define the convention for an arbitrary symmetric, N-player game.

Definition 17. Let $\mathscr{G}$ be a repeated game. For any history $h^t \in \mathscr{H}$, we define the *continuation*

game to be the repeated game that begins in round t following the history h^t.

Definition 18. For a strategy vector $\chi = (\chi_1, \chi_2, \ldots, \chi_N)$ of the repeated game $\mathscr{G}$, we define the *continuation strategy* induced by the history h^t, denoted as $\chi|_{h^t}$, as

$$\chi|_{h^t}(h^\tau) := \chi(h^t h^\tau), \forall h^\tau \in \mathscr{H},$$

where $h^t h^\tau$ is a concatenation of histories h^t and h^τ.

Definition 19. Let $G = (\mathscr{N}, \mathscr{A}, u)$ be a symmetric normal form game and let $\mathscr{G}$ be the repeated version of game G. We define a *convention* as a function ξ that maps each pure-strategy Nash equilibrium $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ of the game G to a continuation strategy vector χ of the repeated game $\mathscr{G}$, such that for any permutation $\eta : \{1,2,\ldots,N\} \leftrightarrow \{1,2,\ldots,N\}$ of the set of players,

$$\big(\xi_1(a_{\eta(1)}, \ldots, a_{\eta(N)}), \ldots, \xi_N(a_{\eta(1)}, \ldots, a_{\eta(N)})\big) = \big(\xi_{\eta(1)}(a_1, \ldots, a_N), \ldots, \xi_{\eta(N)}(a_1, \ldots, a_N)\big) \quad (2.10)$$

that is, *"the convention of a permutation is a permutation of a convention"* (here ξ_i denotes the continuation strategy for player i).

We will use the notation $\eta(\mathbf{a}) := \big(a_{\eta(1)}, \ldots, a_{\eta(N)}\big)$, and $\eta(\xi(\mathbf{a})) := \big(\xi_{\eta(1)}(\mathbf{a}), \ldots, \xi_{\eta(N)}(\mathbf{a})\big)$ to denote the permutation of the history vector using η, and the permutation of the continuation strategy vector respectively.

When two players play the same actions in the pure-strategy NE, the convention assigns both of them the same continuation strategies, as evidenced by the following lemma:

Lemma 8. *Let ξ be a convention, and $\mathbf{a}$ be a pure-strategy NE such that for some players $i, j \in \{1, \ldots, N\}$, $a_i = a_j$. Then*

$$\xi_i(\mathbf{a}) = \xi_j(\mathbf{a}),$$

that is the convention prescribes the same continuation strategies to players i and j.

Proof. Let $\mathbf{a}$ be a pure-strategy NE such that for some players $i, j \in \{1, \ldots, N\}$, $a_i = a_j$.

Define a permutation η as follows:

$$\eta(k) = \begin{cases} j & k = i \\ i & k = j \\ k & \text{otherwise} \end{cases} \quad (2.11)$$

Since $a_i = a_j$, $\mathbf{a} = \eta(\mathbf{a})$. Because ξ is a well-defined function, then

$$\xi(\mathbf{a}) = \xi(\eta(\mathbf{a})). \quad (2.12)$$

From the Definition 19, we know that $\xi(\eta(\mathbf{a})) = \eta(\xi(\mathbf{a}))$. Therefore

$$\xi(\mathbf{a}) = \eta(\xi(\mathbf{a})), \tag{2.13}$$

and $\xi_i(\mathbf{a}) = \xi_{\eta(i)}(\mathbf{a}) = \xi_j(\mathbf{a})$. □

We will formally define the asynchrony round:

Definition 20. Let $\mathcal{G}$ be a repeated symmetric game. We call its round $t_0 \geq 0$ *asynchrony* if in round t_0, the players play some pure-strategy NE of the corresponding stage game, and for all $0 \leq t' < t_0$, the players didn't play a pure-strategy NE in round t'.

To start following a convention, the players have to first learn to play some pure strategy NE. We will call the learning algorithm they will use an *implementation* of a convention. We only limit the definition of the implementation to the resource allocation games.

The implementation maps the outcome from the last round of the game to a strategy for the next round. The mapping is symmetrical. Once an agent becomes a winner and claims some resource, according to the implementation it will keep accessing that resource until the agents reach a pure-strategy NE.

Definition 21. Let $\mathcal{G}_{N,C}$ be an infinitely repeated resource allocation game with N agents and C resources, and let ξ be a convention defined for this game. An *implementation* of a convention ξ is a function π,

$$\pi : \mathcal{A}_1 \times \ldots \times \mathcal{A}_N \cup \{\emptyset\} \to \Delta(\mathcal{A}_1) \times \ldots \times \Delta(\mathcal{A}_N),$$

that is symmetrical (*"the permutation of an implementation is an implementation of a permutation"*). Here, $\pi(\emptyset)$ denotes the strategy vector that the players will adopt in the first round of the game, and $\pi(a_1, a_2, \ldots, a_N)$ denotes the strategy adopted after the players played actions $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ in the last round. Also, for every player i for whom $a_i = A_r$, and for all players $j \neq i$, $a_j \neq A_r$, the strategy χ_i prescribes to play action A_r again, and no other agent will play action A_r with positive probability.

In Chapter 4, we will be concerned with rational strategies for the resource allocation game. That is, we will look its symmetric subgame-perfect equilibria. To construct such equilibria, we define the concepts of an equilibrium convention, and its equilibrium implementation.

Definition 22. An *equilibrium convention* is a convention ξ^* that assigns to any pure-strategy NE of the resource allocation game a continuation strategy that is a subgame-perfect equilibrium of the continuation game.

Definition 23. An *equilibrium implementation* of an equilibrium convention ξ^* is an implementation π^* such that for any action vector $\mathbf{a} \in \mathcal{A}_1 \times \ldots \times \mathcal{A}_N \cup \{\emptyset\}$, the strategy prescribed by the implementation is a best-response to the implementation strategies of the other agents;

that is, the future expected discounted payoff when the agebt adopts implementation π^* and then convention ξ^* is greater or equal to a future expected discount payoff of any other continuation strategy χ'.

Lemma 9. *Let $\mathcal{G}_{N,C}$ be an infinitely repeated resource allocation game. Let ξ^* be an equilibrium convention, and let π^* be its equilibrium implementation. Define a strategy χ_i^* for the agent i as follows:*

1. *Start by following the implementation π^**
2. *When you play a pure-strategy NE of the stage game in round t, start following the convention ξ^* from then on.*

Then the strategy vector χ^* is a symmetric subgame-perfect equilibrium of the infinitely repeated resource allocation game $\mathcal{G}_{N,C}$.

Proof. To show that a given strategy vector χ^* is a subgame-perfect equilibrium, we have to show that for any history of the repeated game h^t, the players are playing the best-response strategy given the strategies of the others.

Let h^t be a history of the game. If there exists $t' < t$ such that in round t', the players played a pure-strategy NE, it means that we will follow a convention ξ^*. Since the convention is a subgame-perfect equilibrium of the continuation game from round t', it means that the players will play indeed their best-response strategies.

Now if for all $t' < t$, the players don't play a pure-strategy NE in round t', they are following the implementation π^*. Therefore, from the definition of the equilibrium implementation, they will play their best-response strategies as well.

By definition the convention and the implementation are symmetric mapping (see Lemma 8), so the resulting strategy vector is also symmetric. □

Figure 2.3 shows how the agents learn to follow a convention when $N = 4$ and $C = 3$. Assume that the players adopt a convention ξ, and they use its implementation π. Initially, they are all "losers", and the implementation prescribes the same strategy to all of them. Once an agent accesses some resource alone, she becomes a winner and will access the same resource until the agents reach an asynchrony round (a state where each resource is accessed by exactly one agent).

2.4 Price of anonymity

In Section 2.1, we have seen that in the stage game of Chicken, the symmetric equilibrium leads to a significantly lower payoff than the asymmetric equilibria. Symmetry of the equilibria

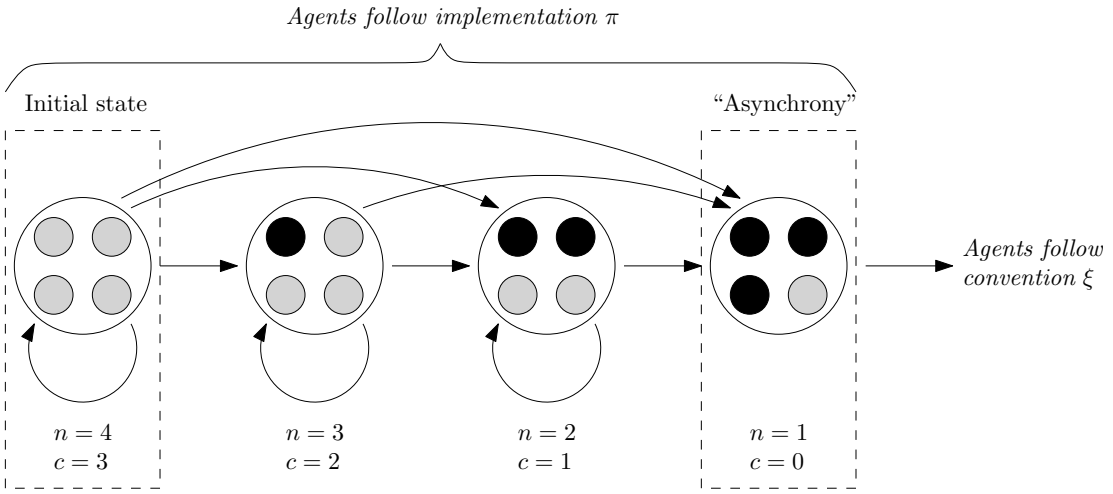


Figure 2.3: Learning to play a convention in a resource allocation game with $N = 4$ agents and $C = 3$ resources. Under each state, we denote the number of losers in the current state n, and the number of unclaimed resources c. Winners are denoted as black circles, losers as light grey circles. In the asynchrony state, there are 3 winners and one loser. Arrows indicate the possible transitions between the states. Once the players reach the asynchrony state, they start following the convention from the next round on.

is a natural requirement when players are all the same, i.e. anonymous. How much social payoff do we have to sacrifice for the requirement of symmetry? Inspired by the price of anarchy of Koutsoupias and Papadimitriou (1999), we propose the *price of anonymity* as a measure of how efficient a given symmetric strategy vector is. For a given symmetric strategy vector of the stage game σ, we calculate the ratio between the social payoff of the most efficient (asymmetric) outcome of the game. The formal definition is as follows:

Definition 24. Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ be a symmetric strategy vector (not necessarily an equilibrium) for the symmetric normal form game $G = (\mathcal{N}, \mathcal{A}, u)$. We define the *price of anonymity of strategy vector* σ as follows:

$$R_G(\sigma) := \frac{\max E(\tau)}{E(\sigma)}$$

where $E(\sigma)$ is the social expected payoff when players adopt strategy σ, and $\max E(\tau)$ is the maximum social payoff of any strategy vector, symmetric or asymmetric.

For a given symmetric game G, we can also define its *price of anonymity* as

$$R_G := \inf R_G(\sigma)$$

where we minimize over all symmetric strategy profiles for the given game. Similarly, we can define the price of anonymity for a symmetric strategy profile χ of the repeated game $\mathcal{G}$, as well as the repeated game $\mathcal{G}$ itself.

As an example, we can take a look at the price of anonymity of the stage game of Chicken from Figure 2.2. The maximum social payoff of 1 is achieved when one player plays *access* and the other player plays *yield*. The symmetric equilibrium with the maximum social payoff is the mixed-strategy Nash equilibrium whose payoff is 0. The price of anonymity of the stage game of Chicken is therefore ∞.

As an example of a strategy profile with finite price of anonymity, consider the infinitely repeated version of the game of Chicken from Figure 2.2. Assume that the players discount the future payoffs with a common discount factor $0 < \delta < 1$. Any strategy profile where the players play one of the two pure strategy Nash equilibria of the Chicken game achieve the social payoff of 1 in each round of the game, so the total discounted social payoff of the most efficient asymmetric strategy is $\frac{1}{1-\delta}$.

Now let the players adopt some convention ξ. In order to implement it, they will start by playing action A randomly with probability p. As soon as the two players take different actions each (and therefore play a pure-strategy NE of the Chicken game), they will start following convention ξ. When the players start following convention ξ, the expected payoff to the winner (the player who accessed alone first) is w_ξ; the expected payoff to the loser (the other player) is l_ξ.

In the randomization rounds, when the other player plays A with probability p, the expected payoff of playing action A is

$$E_A = (1-p)\cdot w_\xi + p\cdot(-2+\delta\cdot E_A),$$

because with probability $1-p$, the player becomes the winner, and with probability p, there will be a collision and the players will face the same situation in the next round. The expected payoff of playing action Y is

$$E_Y = p\cdot l_\xi + (1-p)\cdot\delta\cdot E_Y,$$

because the player becomes the loser with probability p, and with probability $1-p$ none of the players will access, and they will face the same situation in the next round.

In an equilibrium, we want the players to be indifferent between actions A and Y. Therefore, we want $E_A = E_Y$.

When the players adopt the bourgeois convention, the winner payoff is $w_\xi = \frac{1}{1-\delta}$ and the loser payoff is $l_\xi = 0$. Therefore, for any p, $E_Y = 0$ and in the equilibrium, $E_A = E_Y = 0$. The price of anonymity of the bourgeois convention for the Chicken game is infinite. For the egalitarian convention where the players alternate their actions, the winner payoff is $w_\xi = \frac{1}{1-\delta^2}$ and the loser payoff is $l_\xi = \frac{\delta}{1-\delta^2}$. Bhaskar (2000) shows that the convention with the highest possible equilibrium payoff is one where the winner and loser payoff is equal, $w_\xi = l_\xi = \frac{0.5}{1-\delta}$. Such a convention only exists if $\delta \geq 0.5$. We will call such convention the *zero-conflict* convention.

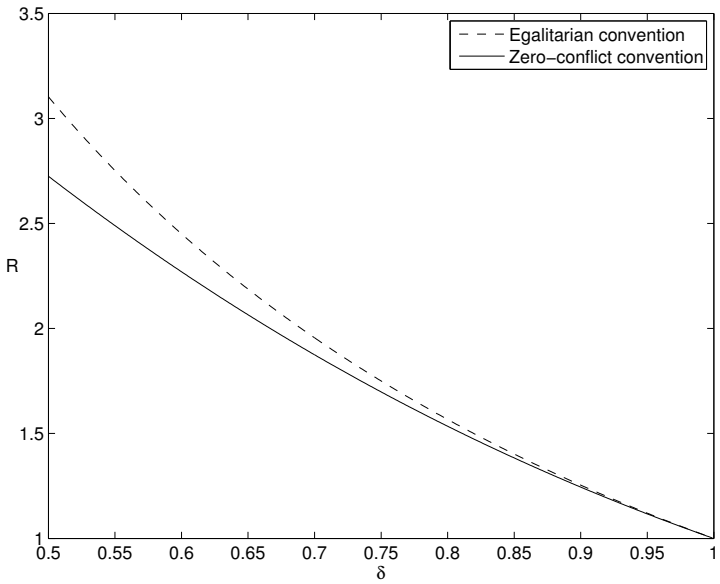


Figure 2.4: Price of anonymity for the equilibrium implementation of the egalitarian and the zero-conflict conventions in the Chicken game.

Solving the system of equations for the arbitrary winner and loser payoffs is difficult. Therefore, we solve the system numerically for the egalitarian convention and for the zero-conflict convention. Figure 2.4 shows the price of anonymity for both conventions, depending on the discount factor δ. For both conventions, the price of anonymity is finite, and it is decreasing as δ increases. In the limit as $\delta \to 1$, the price of anonymity goes to 1. This is because the higher the discount factor, the lower the cost the players pay for the finite learning period, and the higher the relative payoff of the infinite sequence where they follow the convention.

3 Cooperative Resource Allocation

In this chapter, we present a multi-agent learning algorithm for resource allocation in the cooperative setting. Assume that N agents can access C resources in slots (i.e. time intervals of an equal length). In each slot, the agents face an instance of the resource allocation game described in Definition 16 in Section 2.3.2. We assume that after each slot, the agents receive only binary feedback: If they access some resource, they learn whether the access was successful or whether there was a collision; if they yield, they can choose to observe some resource, and they receive information whether the observed resource was free or not.

In Section 2.3.2, we have shown that the resource allocation game has a set of pure-strategy Nash equilibria that correspond to efficient allocations of the C resources to a subset of C agents. The game also has a mixed-strategy Nash equilibrium, where the agents choose to access all the resources with an identical probability. The PSNE are efficient, but not fair, since only C agents can ever access some resource. On the other hand, the MSNE is fair, but not efficient, because the expected social payoff in the MSNE is zero.

We have mentioned the notion of correlated equilibria in Section 2.1. We have shown that for the game of Chicken, there exists a correlated equilibrium that is efficient and fair – a correlation device samples from the set of pure-strategy Nash equilibria of the game, and recommends each agent which action she should play, so that in total the agents play a PSNE. We can use the same principle to find the efficient and fair correlated equilibrium for the general N-agent, C-resource allocation game.

However, the canonical definition of the correlated equilibrium requires a global coordination device that is able to give different recommendations to different agents. Previous work on multi-agent learning (Foster and Vohra (1997); Hart and Mas-Colell (2000); Blum and Mansour (2007)) has presented generic algorithms where the agents learn to play their actions through repeated play. They are then guaranteed to converge to an action profile distribution that is close to a correlated equilibrium. However, these generic algorithms have two significant drawbacks. First, they require additional information (i.e. the agents need to be able to observe the actions played by all the opponents). Second, they do not guarantee a convergence to a

specified efficient and fair correlated equilibrium.

To reach an efficient and fair resource allocation, in this chapter we propose to use a simple signal that all agents can observe at the beginning of each slot and that ergodically fluctuates. This signal could be a common clock, radio broadcasting on a specified frequency, the decimal part of a price of a certain stock at a given time, etc. Depending on this "stupid" signal, "smart" agents can then learn to take a different action for each of its value.

In Section 3.1, we present the algorithm agents use to learn an action for each possible correlation signal value. In Section 3.2 we prove that such an algorithm converges to an efficient correlated equilibrium in polynomial time in the number of agents and resources. We define our measure of fairness, the *Jain index* in Section 3.3, and we show that the fairness of the resulting equilibria (as measured by Jain index) increases as the number of signals K increases. We show that the fairness of the resulting equilibria increases as the number of signals K increases in Section 3.3. Section 3.4 highlights experiments that show the actual convergence rate and fairness. We show how the algorithm performs in case the population is changing dynamically. We also compare the performance of our learning algorithm to generic multi-agent learning algorithms that have been described in the literature. In Section 3.5 we present more related work from game theory and cognitive radio literature, and Section 3.6 concludes.

3.1 Learning Algorithm

In this section, we describe the algorithm that the agents use to learn a correlated equilibrium of the resource allocation game.

Let us denote the space of available coordination signals $\mathscr{K} := \{0,1,\ldots,K-1\}$, and the space of available resources $\mathscr{C} := \{1,2,\ldots,C\}$. Assume that $C \leq N$, that is there are more agents than resources (the opposite case is easier). An agent i has a strategy $f_i : \mathscr{K} \rightarrow \{0\} \cup \mathscr{C}$ that it uses to decide which resource it will access at time t when it receives a correlation signal k_t . When $f_i(k_t) = 0$, the agent does not access at all for signal k_t. The agent stores its strategy simply as a table.

Each agent adapts her strategy as follows:

1. In the beginning, for each $k \in \mathscr{K}$, $f_i(k)$ is initialized uniformly at random from $\mathscr{C}$. That is, every agent picks a random resource to access, and no agent will monitor other resources.
2. At time t, the agent observes signal $k_t \in \mathscr{K}$:
   - If $f_i(k_t) > 0$, the agent tries to access resource $f_i(k_t)$.
   - If otherwise $f_i(k_t) = 0$, the agent chooses a random resource $m_i(t) \in \mathscr{C}$ that it will monitor for activity.

3. Subsequently, the agent observes the outcome of its choice: if the agent accessed some resource, she observes whether the access was successful. If it was, the agent will keep her strategy unchanged. If a collision occurred, the agent sets $f_i(k_t) := 0$ with probability p (with probability $1-p$, the strategy $f_i(k_t)$ stays the same).

4. If the agent did not access, it observes whether the resource $m_i(t)$ it monitored was free. If that resource was free, the agent sets $f_i(k_t) := m_i(t)$ with probability 1.

We can describe this learning algorithm using the notion of convention and its implementation introduced in Chapter 2. For each coordination signal, the agents decide independently, so we only have to define the corresponding convention ξ and implementation π for one coordination signal value only.

The agents who successfully accessed some resource will keep accessing the same resource forever after. Therefore, the agents follow a *bourgeois* convention ξ (defined in Section 2.3), that maps any pure-strategy Nash equilibrium to a continuation strategy vector where the agents play that pure-strategy NE in every round. The implementation π will map an action vector $(a_1, \ldots, a_N)$ to a stage game strategy vector as follows:

- In the first round, each agent chooses action A_r uniformly at random.
- An agent who has played action A_r (access resource r) and didn't collide will play action A_r again.
- An agent who has played action A_r but collided will play action Y with probability p, and action A_r with probability $1-p$.
- An agent who has played action Y will pick a resource r and if no other agent played action A_r in the last round, it will play action A_r in the next round. Otherwise, it will play Y.

Note that since we are not interested in rational strategies, we don't make any claims as to whether the prescribed convention ξ and implementation π is an equilibrium.

3.2 Convergence

An important property of the learning algorithm is if, and how fast it can converge to a pure-strategy Nash equilibrium of the resource allocation game for every signal value. The algorithm is randomized. Therefore, instead of analyzing its worst-case behavior (that may be arbitrarily bad), we will analyze its expected number of steps before convergence.

3.2.1 Convergence for $C = 1, K = 1$

For single resource and single coordination signal, we prove the following theorem:

Theorem 10. *For N agents and $C = 1, K = 1, 0 < p < 1$, the expected number of steps before the allocation algorithm converges to a pure-strategy Nash equilibrium of the resource allocation game is $O\left(\frac{1}{p(1-p)} \log N\right)$.*

To prove the convergence of the algorithm, it is useful to describe its execution as a Markov chain (Definition 14).

When N agents compete for a single signal value, a state of the Markov chain is a vector from $\{0,1\}^N$ that denotes which agents are attempting to access. For the purpose of the convergence proof, it is only important how *many* agents are trying to access, not which agents. This is because the probability with which the agents back-off is the same for everyone. Therefore, we can describe the algorithm execution using the following chain:

Definition 25. A Markov chain describing the execution of the allocation algorithm for $C = 1, K = 1, 0 < p < 1$ is a chain whose state at time t is $X_t \in \{0, 1, \ldots, N\}$, where $X_t = j$ means that j agents are trying to access at time t.

The transition probabilities of this chain look as follows:

$$\Pr(X_{t+1} = N | X_t = 0) = 1 \qquad \text{(restart)}$$
$$\Pr(X_{t+1} = 1 | X_t = 1) = 1 \qquad \text{(absorbing)}$$
$$\Pr\left(X_{t+1} = j | X_t = i\right) = \binom{i}{j} p^{i-j}(1-p)^j \qquad i > 1, j \leq i$$

All the other transition probabilities are 0. This is because when there are some agents accessing a resource, no other agent will attempt to access it.

We are interested in the number of steps it will take this Markov chain to first arrive at state $X_t = 1$ given that it started in state $X_0 = N$ (that is, all the agents are accessing the resource initially). This would mean that the agents converged to a setting where only one of them is accessing, and the others are not. This quantity is known as the *hitting time* (Definition 15).

In Chapter 2, we have shown the Theorem 7 that allows the calculation of the hitting time for Markov chains that satisfy its assumptions. The Markov chain of our algorithm does not satisfy these assumptions though. The problem is that the absorbing state is state 1, and from state 0 the chain goes back to N.

Nevertheless, we can use Theorem 7 to prove the following lemma:

Lemma 11. *Let $A = \{0, 1\}$. The expected hitting time of the set of states A in the Markov chain described in Definition 25 is $O\left(\frac{1}{p} \log N\right)$.*

Proof. We will first prove that the expected hitting time of a set $A' = \{0\}$ in a slightly modified Markov chain is $O\left(\frac{1}{p} \log N\right)$.

Let us define a new Markov chain $(Y_t)_{t\geq 0}$ with the following access probabilities:

$$\Pr(Y_{t+1} = 0 | Y_t = 0) = 1 \qquad \text{(absorbing)}$$
$$\Pr\left(Y_{t+1} = j | Y_t = i\right) = \binom{i}{j} p^{i-j}(1-p)^j \qquad j \geq 0, i \geq 1$$

Note that the access probabilities are the same as in the chain $(X_t)_{t\geq 0}$, except for states 0 and 1. From state 1 there is a positive probability of going into state 0, and state 0 is now absorbing. Clearly, the expected hitting time of the set $A' = \{0\}$ in the new chain is an upper bound on the expected hitting time of set $A = \{0,1\}$ in the old chain. This is because any path that leads into state 0 in the new chain either does not go through state 1 (so it happened with the same probability in the old chain), or goes through state 1, so in the old chain it would stop in state 1 (but it would be one step shorter).

If the chain is in state $Y_t = i$, the next state Y_{t+1} is drawn from a binomial distribution with parameters $(i, 1-p)$. The expected next state is therefore

$$E(Y_{t+1} | Y_t = i) = i(1-p)$$

We can therefore use the Theorem 7 with $\beta := \frac{1}{1-p}$ to derive that for $A' = \{0\}$, the hitting time is:

$$k_i^{A'} < \left\lceil \log_{\frac{1}{1-p}} i \right\rceil + \frac{1}{p} \approx O\left(\frac{1}{p}\log i\right)$$

that is also an upper bound on k_i^A for $A = \{0,1\}$ in the old chain. □

Lemma 12. *The probability h_i that the Markov chain defined in Definition 25 enters state* 1 *before entering state* 0*, when started in any state $i > 1$, is greater than* $1 - p$*.*

Proof. Calculating the probability that the chain X enters state 1 before state 0 is equal to calculating the *hitting probability*, i.e. the probability that the chain ever enters a given state, for a modified Markov chain where the probability of staying in state 0 is $\Pr(X_{t+1} = 0 | X_t = 0) = 1$. For a set of states A, let us denote h_i^A the probability that the Markov chain starting in state i ever enters some state in A. To calculate this probability, we can use Theorem 6. For the modified Markov chain that cannot leave neither state 0 nor state 1, computing h_i^A for $A = 1$ is easy, since the matrix of the system of linear equations is lower triangular.

We'll show that $h_i \geq q = 1 - p$ for $i > 1$ using induction. The first step is calculating h_i for $i \in \{0,1,2\}$.

$$
\begin{aligned}
h_0 &= 0 \\
h_1 &= 1 \\
h_2 &= (1-p)^2 h_2 + 2p(1-p)h_1 + p^2 h_0 \\
&= \frac{2p(1-p)}{1-(1-p)^2} = \frac{2(1-p)}{2-p} \geq 1-p.
\end{aligned}
$$

Now, in the induction step, derive a bound on h_i by assuming $h_j \geq q = 1-p$ for all $j < i$, $j \geq 2$.

$$
\begin{aligned}
h_i &= \sum_{j=0}^{i} \binom{i}{j} p^{i-j}(1-p)^j h_j \\
&\geq \sum_{j=0}^{i} \binom{i}{j} p^{i-j}(1-p)^j q - i\,p^{i-1}(1-p)(q-h_1) - p^i h_0 \\
&= q - i\,p^{i-1}(1-p)(q-1) \geq q = 1-p.
\end{aligned}
$$

This means that no matter which state $i \geq 2$ the Markov chain starts in, it will enter into state 1 earlier than into state 0 with probability at least $1-p$. □

From Lemma 12, we derive that in the original Markov chain (where stepping into state 0 meant going into state N), the chain takes on average $\frac{1}{1-p}$ passes through all its states before it converges into state 1. We know from Lemma 11 that one pass takes in expectation $O\left(\frac{1}{p}\log N\right)$ steps, so the expected number of steps before reaching state 1 is $O\left(\frac{1}{p(1-p)}\log N\right)$. This concludes the proof of Theorem 10.

3.2.2 Convergence for $C \geq 1, K = 1$

Theorem 13. *For N agents and $C \geq 1, K = 1$, the expected number of steps before the learning algorithm converges to a pure-strategy Nash equilibrium of the resource allocation game is $O\left(C\frac{1}{1-p}\left[\frac{1}{p}\log N + C\right]\right)$.*

Proof. In the beginning, in at least one resource, there can be at most N agents who want to access. It will take on average $O\left(\frac{1}{p}\log N\right)$ steps to get to a state when either 1 or 0 agents access (Lemma 11). We will call this period a *round*.

If all the agents backed off, it will take them on average at most C steps before some of them find an empty resource. We call this period a *break*.

The resource might oscillate between the "round" and "break" periods in parallel, but in the worst case, the whole system will oscillate between these two periods.

For a single resource, it takes on average $O\left(\frac{1}{1-p}\right)$ oscillations between these two periods before there is only one agent who accesses in that resource. For $C \geq 1$, it takes on average $O\left(C\frac{1}{1-p}\right)$ steps between "round" and "break" before all resources have only one agent accessing. Therefore, it will take on average $O\left(C\frac{1}{1-p}\left[\frac{1}{p}\log N + C\right]\right)$ steps before the system converges. □

3.2.3 Convergence for $C \geq 1, K \geq 1$

To show what is the convergence time when $K > 1$, we will use a more general problem. Imagine that there are K identical instances of the same Markov chain. We know that the original Markov chain converges from any initial state to an absorbing state in expected time T. Now imagine a more complex Markov chain: In every step, it selects uniformly at random one of the K instances of the original Markov chain, and executes one step of that instance. What is the time T_{all} before all K instances converge to their absorbing states?

This is an extension of the well-known *Coupon collector's problem* (Feller (1968)). The coupon collector problem is the following: Given K coupons, how many coupons do you expect you need to draw with replacement before having drawn each coupon at least once? Here, we are asking a similar question: Given K Markov chains, suppose that in each round we pick one at random and let it take one step. In how many rounds do all the Markov chains converge to an absorbing state? The following theorem (Gast (2011), Theorem 4) shows an upper bound on the expected number of steps after that all the K instances of the original Markov chain converge:

Theorem 14. *(Gast (2011)) Let there be K instances of the same Markov chain that is known to converge to an absorbing state in expectation in T steps. If we select randomly one Markov chain instance at a time and allow it to perform one step of the chain, it will take on average $E[T_{all}] \leq TK\log K + 2TK + 1$ steps before all K instances converge to their absorbing states.*

For arbitrary $C \geq 1, K \geq 1$, the following theorem follows from Theorems 13 and 14:

Theorem 15. *For N agents and $C \geq 1, K \geq 1, 0 < p < 1$, the expected number of steps before the learning algorithm converges to a pure-strategy Nash equilibrium of the resource allocation game for every $k \in \mathcal{K}$ is*

$$O\left((K\log K + 2K)C\frac{1}{1-p}\left[C + \frac{1}{p}\log N\right] + 1\right).$$

Aumann (1974) showed that any Nash equilibrium is a correlated equilibrium, and any convex combination of correlated equilibria is a correlated equilibrium. We also know that all the pure-strategy Nash equilibria that the algorithm converges to are efficient: there are no collisions, and in every resource for every signal value, some agent accesses. Therefore, we conclude the following:

Theorem 16. *The learning algorithm defined in Section 3.1 converges in expected polynomial*

time (with respect to K, C, $\frac{1}{p}$, $\frac{1}{1-p}$ and $\log N$) to an efficient correlated equilibrium of the resource allocation game.

3.3 Fairness

Agents decide their strategy independently for each value of the coordination signal. For each signal value, all the agents use the same randomized algorithm to learn their strategy. Therefore, every agent has an equal chance that the game converges to an equilibrium that is favorable to her. If the agent can access some resource in the resulting equilibrium for a given signal value, we say that the agent *wins* the resource for that signal value. For C available resources and N agents, an agent wins some resource for a given signal value with probability $\frac{C}{N}$ (since no agent can access in two resources for the same signal value).

We can describe the total number of resources won by an agent i as a random variable X_i. This variable is distributed according to a binomial distribution with parameters $\left(K, \frac{C}{N}\right)$.

As a measure of fairness, we use the *Jain index* definedy by Jain et al. (1984). The advantage of Jain index is that it is continuous, so that a resource allocation that is strictly more fair has higher Jain index (unlike measures which only assign binary values, such as whether at least half of the agents access some resource). Also, Jain index is independent of the population size, unlike measures such as the standard deviation of the agent allocation.

For a random variable X, the Jain index is the following:

$$J(X) = \frac{(E[X])^2}{E[X^2]}$$

When X is distributed according to a binomial distribution with parameters $(K, \frac{C}{N})$, its first and second moments are

$$E[X] = K \cdot \frac{C}{N}$$
$$E\left[X^2\right] = \left(K \cdot \frac{C}{N}\right)^2 + K \cdot \frac{C}{N} \cdot \frac{N-C}{N},$$

so the Jain index is

$$J(X) = \frac{C \cdot K}{C \cdot K + (N-C)}. \tag{3.1}$$

For the Jain index it holds that $0 < J(X) \leq 1$. An allocation is considered fair if $J(X) = 1$.

Theorem 17. *For any C, if $K = \omega\left(\frac{N}{C}\right)$, that is the limit $\lim_{N\to\infty} \frac{N}{C\cdot K} = 0$, then*

$$\lim_{N\to\infty} J(X) = 1,$$

so the allocation becomes fair as N goes to ∞.

Proof. The theorem follows from the fact that

$$\lim_{N\to\infty} J(X) = \lim_{N\to\infty} \frac{C\cdot K}{C\cdot K + (N-C)}$$

For this limit to be equal to 1, we need

$$\lim_{N\to\infty} \frac{N-C}{C\cdot K} = 0$$

that holds exactly when $K = \omega\left(\frac{N}{C}\right)$ (note that we assume that $C \leq N$). □

For practical purposes, we may also need to know how big shall we choose K given C and N. The following theorem shows that:

Theorem 18. *Let $\epsilon > 0$. If*

$$K > \frac{1-\epsilon}{\epsilon}\left(\frac{N}{C} - 1\right),$$

then $J(X) > 1-\epsilon$.

Proof. The theorem follows straightforwardly from Equation 3.1. □

3.4 Experimental Results

In all our experiments, we report average values over 128 runs of the same experiment. Error-bars in the graphs denote the interval which contains the true expected value with probability 95%, provided that the samples follow normal distribution. The error bars are missing either when the graph reports values obtained theoretically (Jain index for the constant back-off scheme) or the confidence interval was too small for the scale of the graph.

3.4.1 Static Player Population

Convergence

First, we are interested in the convergence of our allocation algorithm. From Section 3.2 we know that it is polynomial. How many steps does the algorithm need to converge in practice?

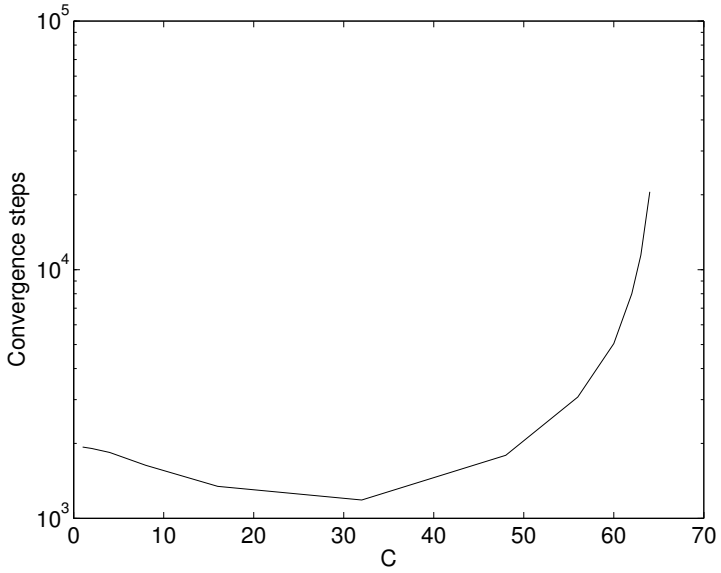


Figure 3.1: Average number of steps to convergence for $N = 64$, $K = N$ and $C \in \{1, 2, \ldots, N\}$.

Figure 3.1 presents the average number of convergence steps for $N = 64$, $K = N$ and increasing number of available resources $C \in \{1, 2, \ldots, N\}$. Interestingly, the convergence takes the longest time when $C = N$. The lowest convergence time is for $C = \frac{N}{2}$, and for $C = 1$ it increases again.

What happens when we change the size of the signal space K? Figure 3.2 shows the average number of steps to convergence for fixed N, C and varying K. Theoretically, we have shown that the number convergence steps is $O(K \log K)$ in Theorem 15. However, in practice the convergence resembles linear dependency on K.

Fairness

From Section 3.3, we know that when $K = \omega\left(\frac{N}{C}\right)$ (that is, K grows asymptotically faster than the ratio $\frac{N}{C}$), the Jain fairness index converges to 1 as N goes to infinity. But how fast is this convergence? How big do we need to choose K, depending on N and C, to achieve a reasonable bound on fairness?

Figure 3.3 shows the Jain index as N increases, for $C = 1$ and $C = \frac{N}{2}$ respectively, for various settings of K. Even though every time when $K = \omega\left(\frac{N}{C}\right)$ (i.e., K grows faster than $\frac{N}{C}$) the Jain index increases, there is a marked difference between the various settings of K. When $K = \frac{N}{C}$, the Jain index is (from Equation 3.1):

$$J(X) = \frac{N}{2N - C}. \tag{3.2}$$

Therefore, for $C = 1$, the Jain index converges to 0.5, and for $C = \frac{N}{2}$, the Jain index is equal to $\frac{2}{3}$ for all $N > 0$, just as Figure 3.3 shows.

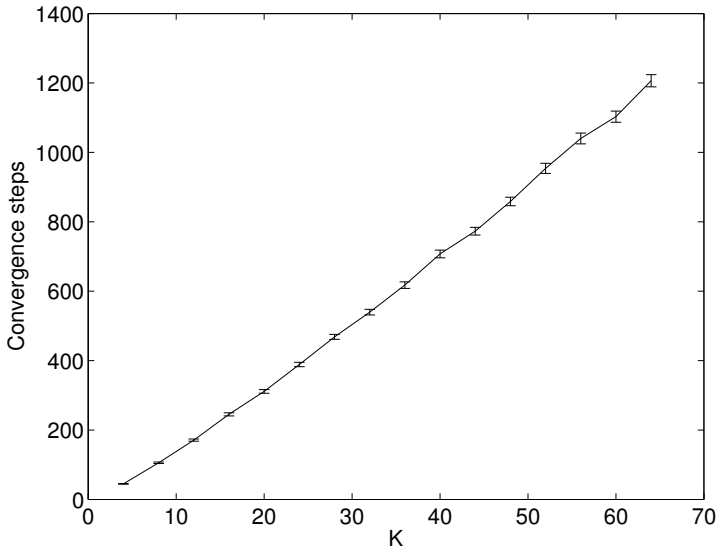
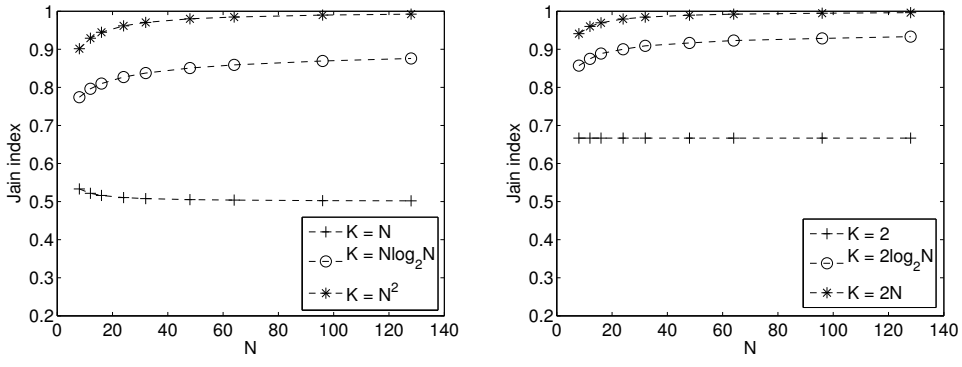


Figure 3.2: Average number of steps to convergence for $N = 64$, $C = \frac{N}{2}$ and $K \in \{2, \ldots, N\}$.



(a) $C = 1$ (b) $C = \frac{N}{2}$

Figure 3.3: Jain fairness index for different settings of C and K, for increasing N.

Optimizing Fairness

We saw how fair the outcome of the allocation algorithm is when agents consider the game for each signal value independently. However, is it the best we can do? Can we further improve the fairness, when each agent correlates her decisions for different signal values?

In a perfectly fair solution, every agent wins some resource for the same number of signal values. However, we assume that agents do not know how many other agents there are in the system. Therefore, the agents do not know what is their fair share of signal values to access in. Nevertheless, they can still use the information for how many signal values they already access some resource to decide whether they should back-off and stop accessing when a collision occurs.

Definition 26. For a strategy f_i^t of an agent i in round t, we define its *cardinality* as the number of signals for which this strategy tells the agent to access:

$$|f_i^t| = \left|\{k \in \mathcal{K} \,|\, f_i^t(k) > 0\}\right|$$

Intuitively, agents whose strategies have at time t higher cardinality should back-off with higher probability than those with a strategy with low cardinality.

We compare the following variations of the resource allocation scheme that differ from the original one only in the probability with which agents back off on collisions:

Constant The scheme described in Section 3.1; Every agent backs off with the same constant probability p.

Linear The back-off probability is $p = \frac{|f_i^t|}{K}$.

Exponential The back-off probability is $p = \mu^{\left(1-\frac{|f_i^t|}{K}\right)}$ for some parameter $0 < \mu < 1$.

Worst-agent-last In case of a collision, the agent who has the *lowest* $|f_i^t|$ does not back off. The others who collided, do back off. This is a greedy algorithm that requires more information than what we assume that the agents have.

To compare the fairness of the allocations in experiments, we need to define the Jain index of an actual allocation. A resource allocation is a vector $\mathbb{X} = (X_1, X_2, \ldots, X_N)$, where X_i is the cardinality of the strategy used by agent i. For an allocation $\mathbb{X}$, its Jain index is:

$$J(\mathbb{X}) = \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N \cdot \sum_{i=1}^{N} X_i^2}$$

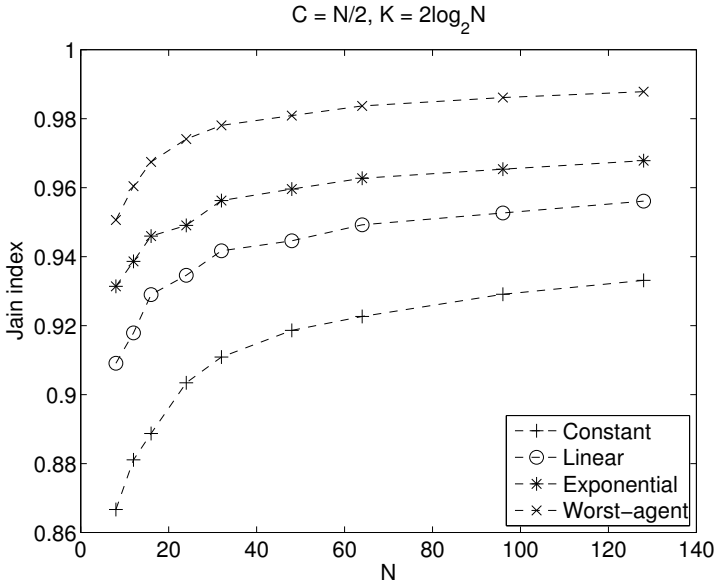


Figure 3.4: Jain fairness index of the resource allocation scheme for various back-off probabilities, $C = \frac{N}{2}$, $K = 2\log_2 N$

Figure 3.4 shows the average Jain fairness index of an allocation for the back-off probability variations. The fairness is approaching 1 for the *worst-agent-last* algorithm. It is the worst if everyone is using the same back-off probability. As the ratio between the back-off probability of the lowest-cardinality agent and the highest-cardinality agent decreases, the fairness increases.

This shows that we can improve fairness by using different back-off probabilities. Nevertheless, the shape of the fairness curve is the same for all of them. Furthermore, the exponential back off probabilities lead to much longer convergence, as shown on Figure 3.5. For $C = \frac{N}{2}$, the convergence time for the linear and constant back-off schemes is similar. The unrealistic *worst-agent-last* scheme is obviously the fastest, since it resolves collisions in 1 step, unlike the other back-off schemes.

3.4.2 Dynamic Player Population

Joining Players

In this section, we will present the results of experiments where a group of players joins the system later. This corresponds to new nodes joining a wireless network. More precisely, 25% of the players join the system from the beginning. The remaining 75% of the players join the system later, one by one. A new player joins the network after the previous players have converged to a perfect resource allocation.

We experiments with two ways of initializing a strategy of a new player.

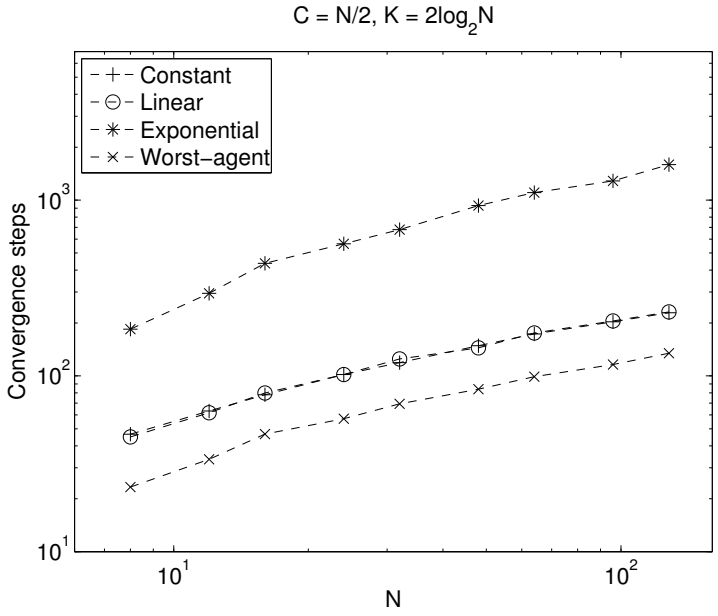


Figure 3.5: Convergence steps for various back-off probabilities.

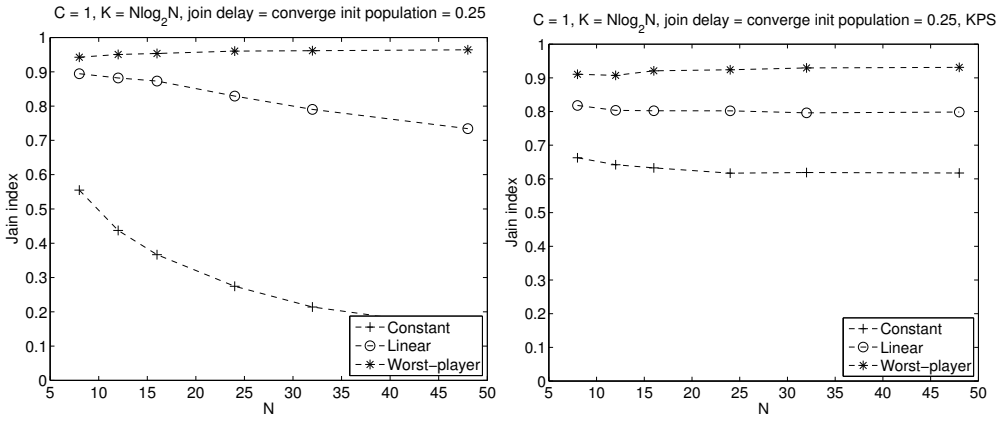
Greedy Either, the joining players cannot observe how many other players there are already in the system. Therefore, their initial strategy tries to access a resource for all possible signal values.

Polite Or, players *do* observe $N(t)$, the number of other players who are already in the system at time t, when the new player joins the system. Therefore, their initial strategy tries to access some resource in for a given signal value only with probability $\frac{1}{N(t)}$.

Figure 3.6 shows the Jain index of the final allocation when 75% of the players join later, for $C = 1$. When the players who join are greedy, they are very aggressive. They start accessing for all signal values. On the other hand, if they are polite, they are not aggressive enough: A new player starts with a strategy that is as aggressive as the strategies of the players who are already in the system. The difference is that the new player will experience a collision for every signal value she accesses in. The old players will only experience a collision in $\frac{1}{N(t)}$ of the signal values for which they access a resource. Therefore, they will back off for less signal values.

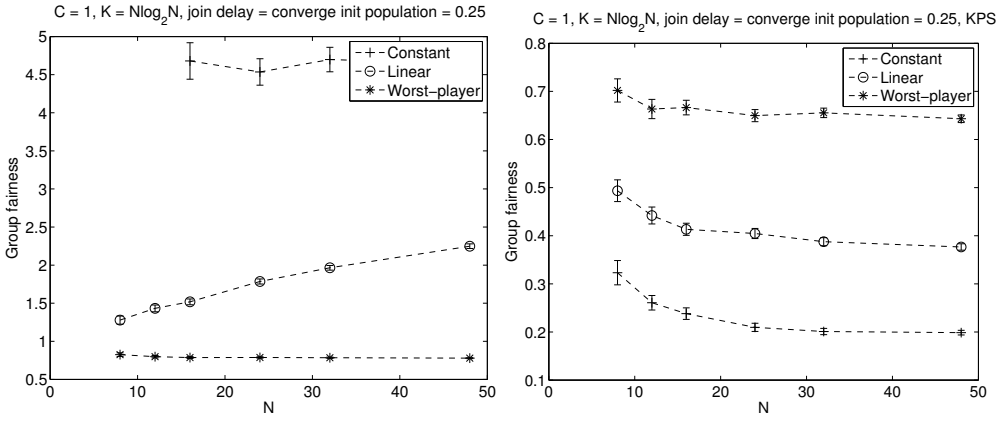
Therefore, especially for the constant scheme, the resulting allocation is very unfair: either it is better for the new players (when they are greedy) or to the older players (when the players are polite).

This phenomenon is illustrated in Figure 3.7. It compares a measure called *group fairness*: the average throughput of the last 25% of players who joined the network at the end ("new" players) divided by the average throughput of the first 25% of players who join the network at the beginning ("old" players).



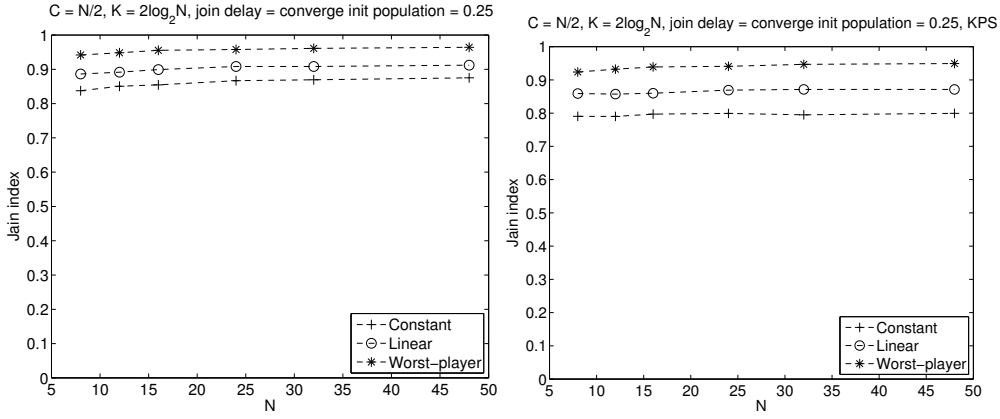
(a) Greedy (b) Polite

Figure 3.6: Joining players, Jain index. $C = 1$ and $K = N\log_2 N$. The two graphs show the results for the two ways of initializing the strategy of a new player.



(a) Greedy (b) Polite

Figure 3.7: Joining players, group fairness. $C = 1$ and $K = N\log_2 N$. The two graphs show the results for the two ways of initializing the strategy of a new player.



(a) Greedy

(b) Polite

Figure 3.8: Joining players, Jain index. $C = \frac{N}{2}$ and $K = 2\log_2 N$. The two graphs show the results for the two ways of initializing the strategy of a new player.

Let's look first at the case when the players are greedy. For the constant scheme, this ratio is around 4.5. For the linear scheme, this ratio is lower, although increasing as N (the total number of players) grows. For the worst-player-last scheme, the ratio stays constant and interestingly, it is lower than 1, which means that "old" players are better off than "new" players.

When players are polite, this situation is opposite. Old players are way better off than new players. For the constant scheme, the throughput ratio is about 0.2.

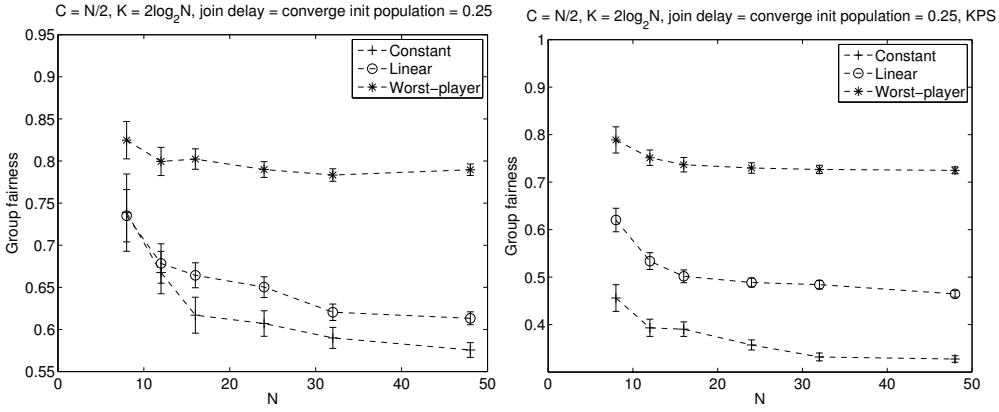
Figures 3.8 and 3.9 show the same graphs for $C = \frac{N}{2}$. Here, the newly joining players are worse off even when they start accessing for every signal value. This is because while they experience a collision every time (because all resources for all signal values are occupied), the old players only experience a collision with a probability $\frac{1}{\frac{N}{2}}$. On the other hand, the overall fairness of the whole population is better, because there are more resources to share and no agent can use more than one resource.

The difference between the old and new players is even more pronounced when the new players are polite.

Restarting Players

Another scenario we looked at was what happens when one of the old players "switches off" and is replaced with a new player with a randomly initialized strategy. We say that such a player got "restarted". Note that the number of players in the network stays the same, it is just that some of the players forget what they have learned and start from scratch.

Specifically, in every round, for every player there is a probability p_R that she will be restarted.



(a) Greedy (b) Polite

Figure 3.9: Joining players, group fairness. $C = \frac{N}{2}$ and $K = 2\log_2 N$. The two graphs show the results for the two ways of initializing the strategy of a new player.

After restart, she will start with a strategy that can be initialized in two ways:

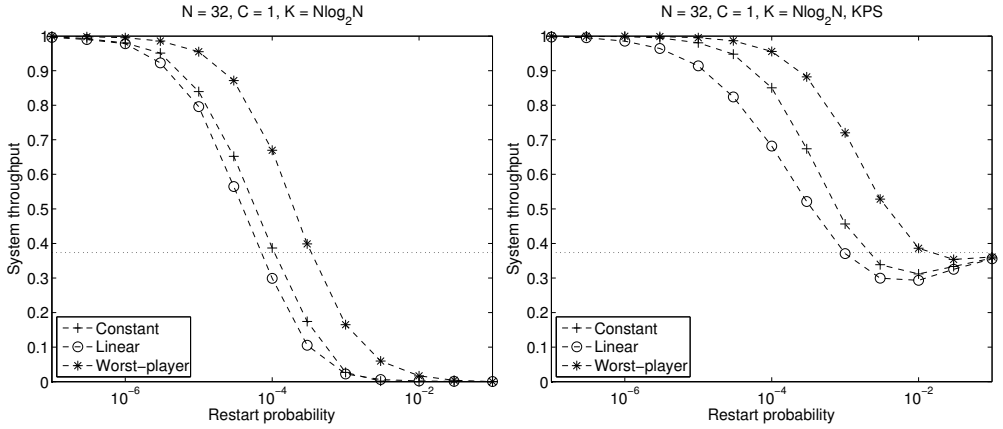
Greedy Assume that the player does not know N, the number of players in the system. Then for each signal value $k \in \mathcal{K}$ she chooses randomly $f_i(k) \in \mathcal{C}$. That means that she attempts to access for every signal value on a randomly chosen resource.

Polite Assume the player *does* know N. For $k \in \mathcal{K}$, she chooses $f_i(k) \in \mathcal{C}$ with probability $\frac{C}{N}$, and $f_i(k) := 0$ otherwise.

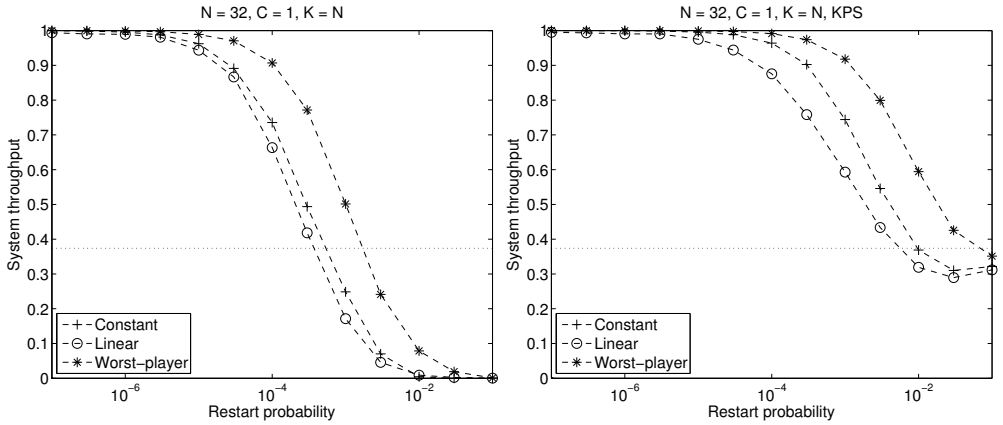
Figure 3.10 shows the average overall throughput when $N = 32$, $C = 1$, and $K = N\log_2 N$ or $K = N$ for the two initialization schemes. A dotted line in all the four graphs shows the overall performance when players attempt to access in a randomly chosen resource with probability $\frac{C}{N}$. This baseline solution reaches $\frac{1}{e} \approx 37\%$ average throughput.

As the probability of restart increases, the average throughput decreases. When players get restarted and they are greedy, they attempt to access for every signal value. If there is only one resource available, this means that such a restarted player causes a collision for every signal value. Therefore, it is not surprising that when the restart probability $p_R = 10^{-1}$ and $N = 32$, the throughput is virtually 0: In every step, in expectation at least one player will get restarted, so there will be a collision almost always.

There is an interesting "phase transition" that occurs when $p_R \approx 10^{-4}$ for $K = N\log_2 N$, and when $p_R \approx 10^{-3}$ for $K = N$. There, the performance is about the same as in the baseline random access scenario (that requires the players to know N though). Similar phase transition occurs when players are polite, even though the resulting throughput is higher, since the restarted players are less "aggressive".

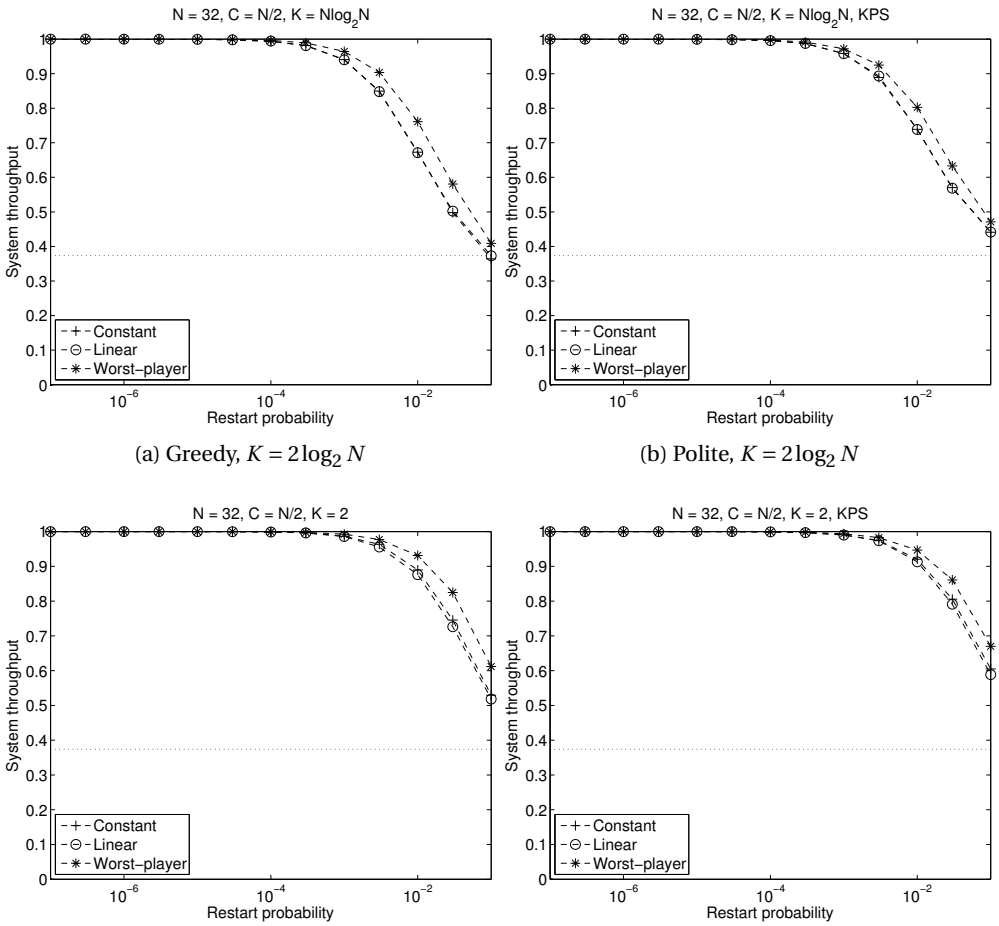


(a) Greedy, $K = N\log_2 N$ (b) Polite, $K = N\log_2 N$



(c) Greedy, $K = N$ (d) Polite, $K = N$

Figure 3.10: Restarting players, throughput, $N = 32$, $C = 1$



(a) Greedy, $K = 2\log_2 N$

(b) Polite, $K = 2\log_2 N$

(c) Greedy, $K = 2$

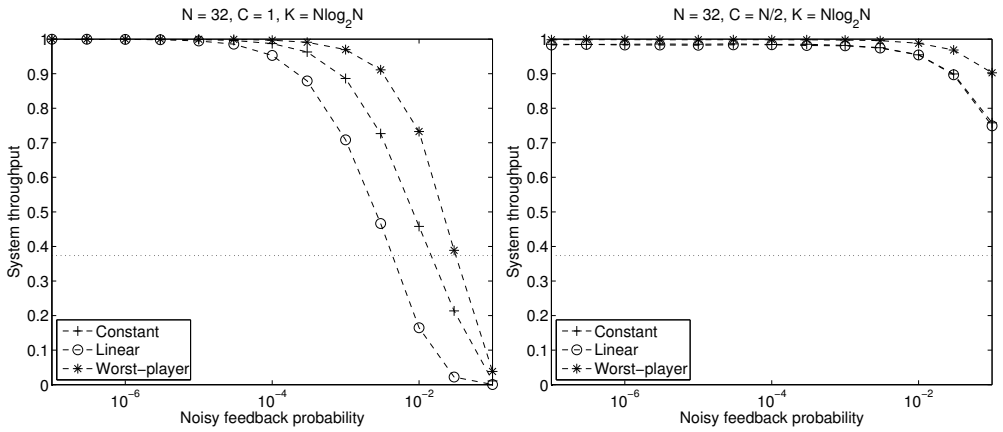
(d) Polite, $K = 2$

Figure 3.11: Restarting players, throughput, $N = 32$, $C = \frac{N}{2}$

Yet another interesting, but not at all surprising, phenomenon is that while the "worst-player-last" scheme still achieves the highest throughput, the "constant" back off scheme is better than the "linear" back-off scheme. This is because for the average overall throughput, it only matters how fast are the players able to reach a perfect allocation after a disruption. The worst-player-last scheme is the fastest, since it resolves a collision in 1 step. The constant scheme with $p_R = \frac{1}{2}$ is worse (see Theorem 15). The linear scheme is the slowest.

Figure 3.11 shows the average overall throughput for $C = \frac{N}{2}$, and $K = \log_2 N$ or $K = 2$. There is no substantial difference between when players are greedy or polite. Since there are so many resources available, a restarted player will only cause a small number of collisions (in one resource out of $\frac{N}{2}$ for every signal value), so the throughput will not decrease too much.

Also, the convergence time for linear and constant scheme is about the same when $C = \frac{N}{2}$, so they both adapt to the disruption equally well.



(a) $C = 1$, $K = N \log_2 N$ (b) $C = \frac{N}{2}$, $K = 2\log_2 N$

Figure 3.12: Noisy feedback, throughput, $N = 32$

Noisy Feedback

So far we assumed that players receive perfect feedback about whether their accesses were successful or not. They could also observe the activity on a given resource perfectly. We are going to loosen this assumption now.

Suppose that in every step, every player has a probability p_F that the feedback she receives was wrong. That is, if the player accessed, she will learn that the access was successful when it was not, and vice versa. If the player observed some resource, she will learn that the resource was free when in fact it was not (and vice versa). In the context of wireless networks, this corresponds to an interference on the wireless channel.

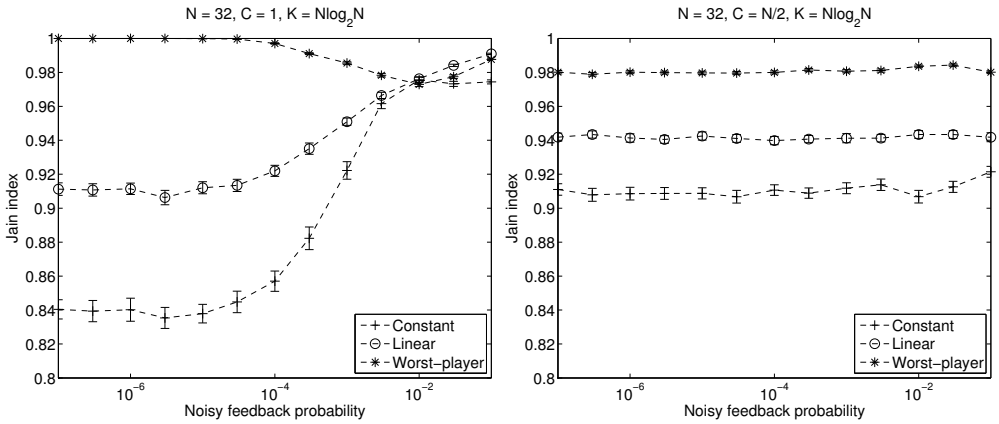
How does this affect the learning?

In Figure 3.12 we show the average overall throughput when $C = 1$ and $C = \frac{N}{2}$ respectively. For one resource, the constant scheme is better than the linear scheme, because it adapts faster to disruptions. For $C = \frac{N}{2}$, both schemes are equivalent, because they are equally fast to adapt. A phase transition occurs when the noisy feedback probability is about $p_F = 10^{-2}$.

Figure 3.13 shows the Jain index of the allocation when players receive noisy feedback. As usual, the linear scheme is better than the constant. Only when the overall throughput drops close to 0, all the schemes obviously have almost the same fairness.

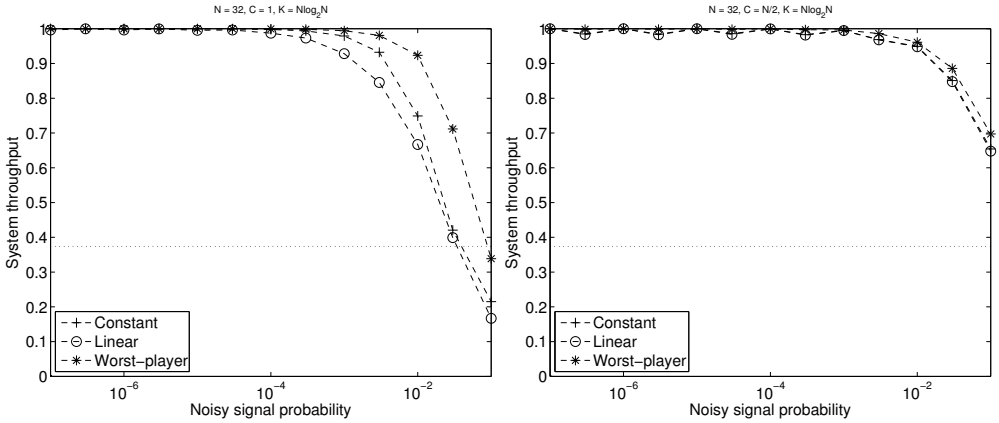
Noisy Coordination Signal

Our algorithm assumes that all players can observe the same coordination signal in every step. But where does this signal come from? It may be some random noise on a given frequency, an FM radio transmission etc. However, the coordination signal might be noisy, and different



(a) $C = 1$, $K = N\log_2 N$ (b) $C = \frac{N}{2}$, $K = 2\log_2 N$

Figure 3.13: Noisy feedback, Jain index, $N = 32$



(a) $C = 1$, $K = N\log_2 N$ (b) $C = \frac{N}{2}$, $K = 2\log_2 N$

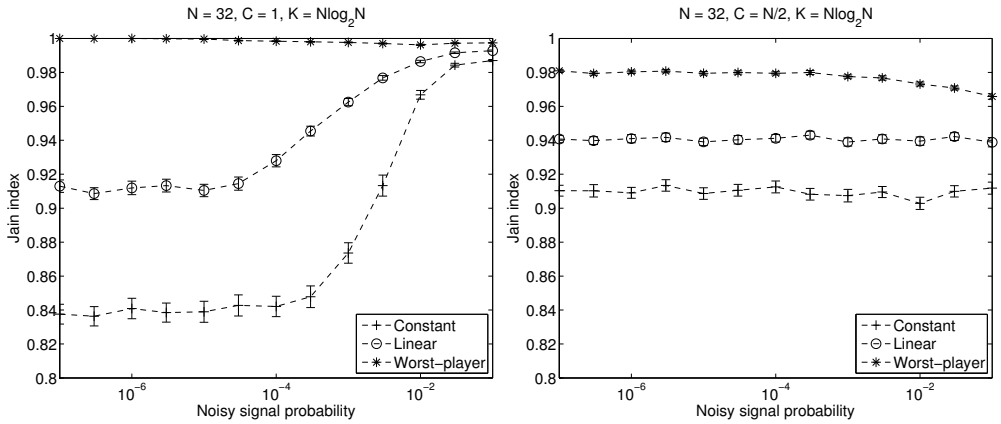
Figure 3.14: Noisy coordination signal, throughput, $N = 32$

players can observe a different value. This means that their learning would be "out of sync". In the wireless networks, this corresponds to *clock drift*.

To see what happens in such a case, we use the following experiment. In every step, every player observes the correct signal (i.e. the one that is observed by everyone else) with probability $1 - p_S$. With probability p_S it observes some other false signal (that is still taken uniformly at random from the set $\{0, ..., K-1\}$).

The overall throughput is shown in Figure 3.14. We can see that the system is able to cope with a fairly high level of noise in the signal, and the drop in throughput only occurs as $p_S = 10^{-1}$.

The Jain index of the allocation (Figure 3.15) stays almost constant, only when the throughput drops the Jain index increases. When the allocation is more random, it is also more fair.



(a) $C = 1$, $K = N \log_2 N$ (b) $C = \frac{N}{2}$, $K = 2 \log_2 N$

Figure 3.15: Noisy coordination signal, Jain index, $N = 32$

3.4.3 Generic Multi-agent Learning Algorithms

Several algorithms that are proved to converge to a correlated equilibrium have been proposed in the multi-agent learning literature. In the Introduction, we have mentioned three such learning algorithms (Foster and Vohra (1997); Hart and Mas-Colell (2000); Blum and Mansour (2007)). However, the analysis of Foster and Vohra was only applicable to games of two players. In this section, we will briefly recall the other two multi-agent learning algorithms (Hart and Mas-Colell (2000); Blum and Mansour (2007)), and compare their performance with our algorithm presented in Section 3.1.

The two algorithms we will compare our algorithm to are based on the notion of minimizing *regret* the agents experience from adopting a certain strategy. Intuitively, we can describe the concept of regret as follows: Imagine that an agent uses strategy σ in a couple of rounds of the game, and accumulates a certain payoff. We would like to know how does this payoff compare to a payoff acquired by some simple alternative strategy τ. The difference in the payoff between the strategy τ and σ is the regret the agent perceives (ex-post) for choosing strategy σ over strategy τ.

What do we mean by "simple strategy"? One class of simple strategies are strategies that always select the same action. The *external regret* compares the performance of the strategy σ to the performance of the best single action ex-post.

Another class of alternative strategies are strategies that modify strategy σ slightly. Every time the strategy σ proposes to play action a, the alternative strategy τ proposes action $a' \neq a$ instead. The *internal regret* is defined as the regret of strategy σ compared to the best such alternative strategy. When all the agents adopt a strategy with low internal regret, they converge to a strategy profile that is close to a correlated equilibrium (also shown by Blum and Mansour).

Hart and Mas-Colell present a simple multi-agent learning algorithm that is guaranteed to converge to a correlated equilibrium. They assume that the players can observe the actions of all their opponents in every round of the game. Players start by choosing their actions randomly. Then they update their strategy as follows: Let a_i be the action that player i played in round $t-1$. For each action $a_j \in \mathcal{A}_i$, $a_j \neq a_i$, player i calculates the difference between the average payoff she would have received had she played action a_j instead of a_i in the past, and the average payoff she received so far while playing action a_i. As we mentioned above, we can call this difference the *internal regret* of playing action a_i instead of action a_j. The player then chooses the action to play in round t with probability proportional to its internal regret compared to the previous action a_i. Actions with negative regret are never played. The previous action a_i is played with positive probability – this way, the strategy has a certain inertia.

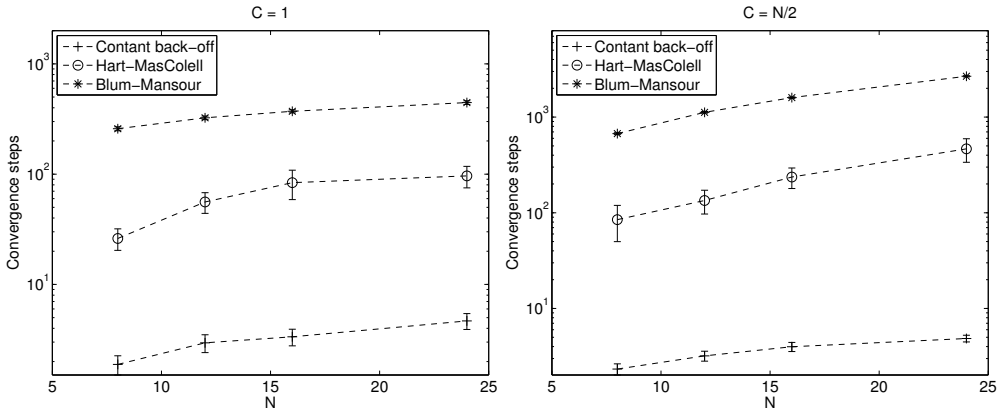
Hart and Mas-Colell prove that if the agents adopt the adaptive procedure described above, the empirical distribution of the play (the relative frequency of playing a certain pure strategy profile) converges almost surely to the set of correlated equilibria.

Blum and Mansour present a general technique to convert any learning algorithm with low external regret to an algorithm with a low internal regret. The idea is to run multiple copies of the external regret algorithm. In each step, each copy returns a probability vector of playing each action. These probability vectors are then combined into one joint probability vector. When the player observes the payoff of playing each action, she updates the payoff beliefs of each external regret algorithms proportionally to the weight they had in the joint probability vector. The authors then show that when the players all use a learning algorithm with low internal regret, the empirical distribution of the game converges close to a correlated equilibrium.

One of the low-external-regret algorithms that Blum and Mansour present is the *Polynomial Weights (PW)* algorithm. There, a player keeps a weight for each of her actions. In every round of the game, she updates the weight proportionally to the loss (negative payoff) that action incurred in that round. Actions with higher weight get then chosen with a higher probability.

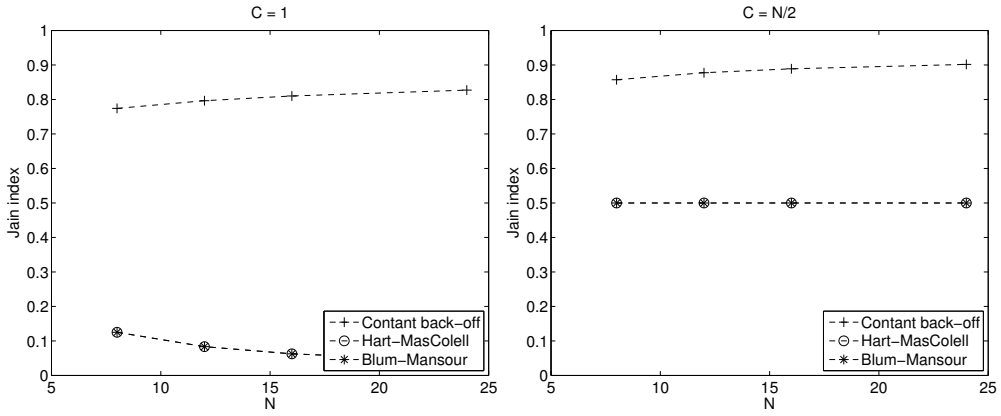
We have implemented both the internal-regret-based algorithm of Hart and Mas-Colell (2000), and the PW algorithm of Blum and Mansour (2007). In all our experiments, both algorithms always converge to a pure-strategy Nash equilibrium of the resource allocation game, and therefore to an efficient allocation. However, the resulting allocation is not fair, as only a subset of agents of size C can ever access the resources.

Figure 3.16 shows the average number of rounds the algorithms take to converge to a stable outcome. We compare their performance with our learning algorithm from Section 3.1. For our learning algorithm, we set $K = 1$, so that it also only converges to a pure-strategy Nash equilibrium of the game. We performed 128 runs of each algorithm for each scenario. The error-bars in Figure 3.16 show the 95% confidence interval of the average, assuming that the convergence times are distributed according to a normal distribution.



(a) $C = 1$ (b) $C = \frac{N}{2}$

Figure 3.16: General multi-agent learning algorithms, convergence rate.



(a) $C = 1$ (b) $C = \frac{N}{2}$

Figure 3.17: General multi-agent learning algorithms, Jain index.

Not surprisingly, the generic algorithms of Hart and Mas-Colell (2000) and Blum and Mansour (2007) cannot match the convergence speed of our algorithm, designed specifically for the problem of resource allocation. As the generic algorithms converge to a pure-strategy NE, the outcome is very unfair, and the Jain index is very low, as evidenced by Figure 3.17. We don't report the confidence bounds for the Jain index, as in all of the experiments the resulting Jain index was the same.

It is worth noting that since the algorithms are randomized, any agent has the same chance of winning a resource. Therefore, if we used the generic algorithms with a coordination signal like the one used by our algorithm, the resulting allocation would be equally as fair as the allocation achieved by our constant back-off algorithm. Nevertheless, the generic algorithms take longer to converge to an efficient allocation than our special purpose algorithm.

3.5 Related Work

Broadly speaking, the resource allocation game we are interested in this thesis belongs to a class of games with the following property: The payoff an agent receives from a certain action is inversely proportional to the number of other agents who chose the same action. How can we achieve efficient and fair outcome in such games, provided that the agents are cooperative and follow our prescribed protocol? Variants of this problem have been studied in several previous works.

The simplest such variant is the *Minority game* (Challet et al. (2005)). In this game, N agents have to simultaneously choose between two actions. Agents who chose an action that was chosen by a minority of agents receive a payoff of 1, whereas agents whose action choice was in majority receive a payoff of 0.

This game has many pure-strategy Nash equilibria, where some group of $\lfloor \frac{N-1}{2} \rfloor$ agents chooses one action and the rest choose the other action. Such equilibria are efficient, since the largest possible number of agents achieve the maximum payoff. However, they are not fair: the payoff to the losing group of agents is always 0. This game has also one mixed-strategy NE that is fair: every agent chooses its action randomly. This equilibrium, on the other hand, is not efficient: the expected size of the minority group is lower than $\lfloor \frac{N-1}{2} \rfloor$ due to variance of the action selection.

Savit et al. (1999) show that if the agents receive feedback on which action was in the minority, they can learn to coordinate better to achieve a more efficient outcome in a repeated minority game. They do this by basing the agents' decisions on the history of past iterations. Cavagna (1999) shows that the same result can be achieved when agents base their decisions on the value of some random coordination signal instead of using the history. This is a direct inspiration for the idea of global coordination signal presented in this chapter.

The ideas from the literature on Minority games have recently found their way into the

cognitive radio literature. Mahonen and Petrova (2008) present a resource allocation problem much like ours. The agents learn which resource they should use using a strategy similar to the strategies for minority games. The difference is that instead of preferring the action chosen by the minority, in the resource allocation problem, an agent prefers resources which were not chosen by anyone else. Using this approach, Mahonen and Petrova are able to achieve a stable throughput of about 50% even when the number of agents who try to access over a resource increases. However, each agent is essentially choosing one out of a fixed set of strategies, that they cannot adapt. Therefore, it is very difficult to achieve a perfectly efficient resource allocation.

Another, more general variant of our problem, called *dispersion game* was described by Grenager et al. (2002). In a dispersion game, agents can choose from several actions, and they prefer the one that was chosen by the smallest number of agents. The authors define a *maximal dispersion outcome* as an outcome where no agent can move to an action with fewer agents. The set of maximal dispersion outcomes corresponds to the set of pure-strategy Nash equilibria of the game. They propose various strategies to converge to a maximal dispersion outcome, with different assumptions on the information available to the agents. On the contrary with our work, the individual agents in the dispersion games do not have any particular preference for the actions chosen or the equilibria which are achieved. Therefore, there are no issues with achieving a fair outcome.

Verbeeck et al. (2007) use reinforcement learning, namely *linear reward-inaction automata*, to learn Nash equilibria in common and conflicting interest games. For the class of conflicting interest games (to which our resource allocation game belongs), they propose an algorithm that allows the agents to circulate between various pure-strategy Nash equilibria, so that the outcome of the game is fair. In contrast with our work, their solution requires more communication between agents, and it requires the agents to *know* when the strategies converged. In addition, linear reward-inaction automata are not guaranteed to converge to a PSNE in conflicting interest games; they may only converge to pure strategies.

All the games discussed above, including the resource allocation game, form part of the family of *potential games* introduced by Monderer and Shapley (1996). A game is called a potential game if it admits a *potential function*. A potential function is defined for every strategy profile, and quantifies the difference in payoffs when an agent unilaterally deviates from a given strategy profile. There are different kinds of potential functions: exact (where the difference in payoffs to the deviating agent corresponds directly to the difference in potential function), ordinal (where just the sign of the potential difference is the same as the sign of the payoff difference) etc.

Potential games have several nice properties. The most important is that any pure-strategy Nash equilibrium is just a local maximum of the potential function. For finite potential games, players can reach these equilibria by unilaterally playing the best-response, no matter what initial strategy profile they start from.

The existence of a natural learning algorithm to reach Nash equilibria makes potential games an interesting candidate for our future research. We would like to see to which kind of correlated equilibria can the agents converge there, and if they can use a simple correlation signal to coordinate.

3.6 Conclusions

In this chapter, we proposed a new approach to reach efficient and fair solutions in cooperative multi-agent resource allocation problems. Instead of using a centralized, "smart" coordination device to compute the allocation, we use a "stupid" coordination signal, in general a random integer $k \in \{0, 1, \ldots, K-1\}$ that has no a priori relation to the problem. Agents then are "smart": they learn, for each value of the coordination signal, which action they should take.

From a game-theoretic perspective, the ideal outcome of the game is a correlated equilibrium. Our results show that using a global coordination signal, agents can learn to play a convex combination of pure-strategy Nash equilibria, that is a correlated equilibrium.

We showed a learning strategy that, for the resource allocation game defined in Chapter 2, converges in expected polynomial number of steps to an efficient correlated equilibrium. We also proved that this equilibrium becomes increasingly fair as K, the number of available synchronization signals, increases.

We have confirmed both the fast convergence as well as increasing fairness with increasing K experimentally. We have also investigated the performance of our learning strategy in case the agent population is dynamic. When new agents join the population, our learning strategy is still able to learn an efficient allocation. However, the fairness of this allocation will depend on how greedy the initial strategies of the new agents are. When agents restart at random intervals, it becomes more important how fast a strategy converges. A simple strategy where everyone backs off from accessing with a constant probability is able to achieve higher throughput than a more sophisticated strategy where the back-off probability depends on for how many signal values an agent is already accessing. Finally, we have shown experimentally that the learning strategy is robust against noise in both the coordination signal, as well as in the feedback the agents receive about resource use.

4 Non-cooperative Resource Allocation

In the previous chapter, we have shown how the agents can use a global coordination signal to reach an efficient and fair resource allocation when the agents are cooperative. However, the proposed allocation algorithm was not rational – a self-interested agent could keep accessing a resource forever, until everyone else backs off.

In this chapter, we will analyze resource allocation protocols that are rational. Specifically, we will assume that the players play an infinitely repeated version of the resource allocation game, and that they discount the future payoffs with a common discount factor $0 < \delta < 1$. In contrast with the previous chapter, we assume full observability – after each round, the agents receive a feedback about occupancy of all the resources. Our goal is to obtain resource allocation protocols that are subgame-perfect equilibria of the infinitely repeated game with discounting. The subgame-perfect equilibria that we look for will all have the following structure: All agents start playing the same randomized strategy. When their actions differ, the will adopt a convention (defined in Chapter 2, Definition 19).

In Section 4.1, we extend the definition of a convention from Section 2.3 to an *augmented* convention, that allows the agents to use a global coordination signal to condition their strategies on. We show that for any equilibrium augmented convention, there exists an equilibrium implementation, such that when the agents play according to the convention and its implementation, they play a subgame-perfect equilibrium of the resource allocation game. In Section 4.2, we present two examples of a convention: the bourgeois convention, and the market convention. We analyze their efficiency and convergence properties. Finally, Section 4.4 concludes.

4.1 Resource Allocation Game

In Chapter 2, we have defined the resource allocation game of N-agents and C resources (Definition 16). We have shown that while it is a symmetric game, the only efficient equilibria of the stage game are asymmetric. The only symmetric Nash equilibrium is the mixed strategy

NE. However, for low number of resources C, this equilibrium has zero expected social payoff.

Therefore we turned our attention to symmetric equilibria of the repeated resource allocation game. We have defined the convention and its implementation (Definition 19 and Definition 21). We have shown that given an equilibrium convention and its equilibrium implementation, we can construct a symmetric subgame-perfect equilibrium of the repeated resource allocation game. In this section, we will show how we can find the equilibrium implementation for a given equilibrium convention.

The convention assigns the agents a continuation strategy depending on which action they play in an asymmetric pure-strategy NE. We can say that the agents get assigned a *role*. In the resource allocation game, we call the agents who access some resource in the pure-strategy NE the "winners". The other agents (those who yield) are called "losers".

The problem with the convention is that it won't distinguish between the losers, since they play the same action, to yield. They will all have to adopt the same continuation strategy. In order to distinguish between them, and to get a richer set of possible conventions, we will adopt the idea from Chapter 3. We will assume that the agents can observe in each round of the game a global coordination signal – an integer $k \in \{1,2,\ldots,K\}$ chosen uniformly at random. The agents then condition their strategy on the coordination signal value. That way, for different signal values, there can be different sets of winners and losers.

We will augment the definitions of a convention to include the coordination signal. Without specifying otherwise, we will use the augmented definitions in this chapter.

Definition 27. Let $\mathscr{G}$ be an infinitely repeated game with discounting. We define the *augmented history* $\hat{h}_t$ of the play in round $t \geq 0$ as

$$\hat{h}_t := \left(((a_1^0, a_2^0, \ldots, a_N^0), k_0), \ldots, ((a_1^{t-1}, a_2^{t-1}, \ldots, a_N^{t-1}), k_{t-1})\right)$$

where a_i^t is the action taken by agent i in round t, and k_t is the signal that the agents observe in round t.

Similarly, we can augment the strategy of the repeated game to take into account the signal as well:

Definition 28. An *augmented strategy in the repeated game* of an agent i is a function from the augmented history and a currently observed coordination signal to a probability distribution over the action space,

$$\hat{\chi}_i : (\hat{h}^t, k_t) \mapsto \Delta(\mathscr{A}_i)$$

Definition 29. Let $G = (\mathscr{N}, \mathscr{A}, u)$ be a symmetric normal form game and let $\mathscr{G}$ be the repeated version of game G. We define an *augmented convention* as a function $\hat{\xi}$ that maps a vector of

Round	**1**	2	**3**	**4**	5	6	**7**	8
Signal	1	2	2	2	1	2	1	2
Agent 1	1	0	2	0	1	0	***1***	0
Agent 2	1	0	***1***	1	1	1	0	1
Agent 3	***2***	1	0	0	2	0	2	0
Agent 4	0	1	2	***2***	1	2	0	2

Figure 4.1: Example of a game play for $N = 4$ agents, $C = 2$ and $K = 2$. Once an agent accesses a resource alone, it will keep accessing that resource every time the same signal is observed. The winners are denoted with grey background. The first round when an agent accesses a resource alone (and becomes the winner) is denoted with bold face. In the rest of the game, agent 1 will keep accessing resource 1 when the signal is 1. Agent 2 will access the resource 1 when signal is 2. Agent 3 will access the resource 2 when the signal is 1. Finally, agent 4 will access the resource 2 when the signal is 2.

pure-strategy Nash equilibria of the game G for each signal value $\mathbf{a} = (\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^K)$ to a vector of augmented continuation strategies of the repeated game $\mathscr{G}$, such that for any permutation $\eta : \{1,2,\ldots,N\} \leftrightarrow \{1,2,\ldots,N\}$ of the set of players,

$$\hat{\xi}((\eta(\mathbf{a}^1),\ldots,\eta(\mathbf{a}^K))) = \eta(\hat{\xi}(\mathbf{a}^1,\mathbf{a}^2,\ldots,\mathbf{a}^K)) \tag{4.1}$$

that is, as before, *"the augmented convention of a permutation is a permutation of an augmented convention"*. The continuation strategies can be different for each coordination signal value.

Figure 4.1 gives an example of a game play of $N = 4$ agents, $C = 2$ resources and $K = 2$ signals. If in round t, the agents observe a signal k_t, the augmented convention adopted by the agents in this example prescribes that if an agent accesses a resource alone in round t, it becomes its "winner" and will access the same resource in every round $t' > t$ where the signal $k_{t'} = k_t$.

Similarly as in Chapter 2, we define an *augmented implementation* $\hat{\pi}$ for an augmented convention $\hat{\xi}$. An augmented implementation maps the last outcomes for each signal, together with the signal the agents observe in the current round, to a strategy for the current round. It has to satisfy the same requirements as the regular implementation: The agents who accessed some resource for the given signal before will access it again. The augmented implementation is also symmetrical, in that two agents who play the same action for all signals have to play the same strategy now.

We will prove an existence of an equilibrium implementation for a set of *uniform* conventions. In a uniform convention, all the winners (that is, all the players who access some resource in the pure-strategy NE) get the same payoff $w_{\hat{\xi}}$. Naturally, all the losers get all the same payoff $l_{\hat{\xi}}$ too.

The implementations we will be looking for have all the same structure. Depending on the number of losers, on the number of free resources, and the current signal value, they assign an

identical mixed strategy to all the losers:

Definition 30. Let $\hat{\xi}$ be a uniform augmented convention. Let k be a coordination signal value, let n be the number of agents who have not yet won a resource for signal k, and let c be the number of unclaimed resources for signal k. A *uniform augmented implementation* of a convention $\hat{\xi}$ is a function $\hat{\pi}$,

$$\hat{\pi} : (n, c, k) \mapsto 0 \leq p \leq 1,$$

such that all the losers play the action A with probability p. All the winners access the same resource as in the previous round of the game.

If in round t there are n losers, c unclaimed resources and the agents observe the signal k, $\hat{\pi}(n, c, k)$ defines a mixed strategy for the losers. When the losers play action A with probability $\hat{\pi}(n, c, k)$, then either

1. $\hat{\pi}(n, c, k) = 1$ and all the losers prefer to play A, or
2. $\hat{\pi}(n, c, k) = 0$ and all the losers prefer to play Y, or
3. $0 < \hat{\pi}(n, c, k) < 1$ and all the losers are indifferent between playing A and Y.

Note that the number of losers only decreases. So either we are in the same situation, or there is a few losers less. We can define the extended convention ξ^* such that it maps any outcome where there are less losers to a continuation strategy.

Definition 31. Let $\hat{\xi}$ be a uniform augmented convention, and $\hat{\pi}$ its uniform augmented implementation. We can define the *extended convention* $\hat{\xi}'$ as a function

$$\hat{\xi}' : (n, c, k) \mapsto (\hat{\chi}_w, \hat{\chi}_l),$$

where $\hat{\chi}_w$ is the continuation strategy for the new winners and $\hat{\chi}_l$ is the continuation strategy for the losers.

It can be easily shown that if the convention ξ is uniform, the extended convention is uniform as well.

Definition 32. Let $\hat{\xi}$ be a uniform augmented convention for the resource allocation game $\mathcal{G}_{N,C}$. In round t, let there be n losers, and c unclaimed resources. Let $\hat{\chi}$ be a pure strategy of an agent α who is a loser too. Assume that for each signal $k \in \{1, \ldots, K\}$, every other loser takes action A with probability p_k. Let $\mathbf{p} = (p_1, p_2, \ldots, p_K)$ be a vector of these probabilities, that is, the strategy of the other losers. We define *expected payoff functions* E_A *and* E_Y when agent α

takes actions A and Y :

$$E_A(\mathbf{p},\hat{\chi},k) := \sum_{n_w=1}^{\min(n,c)} \Big[\Pr(\alpha \text{ wins } \& \ n_w \text{ winners}|A)\, w_{\hat{\xi}}(n_w) + \Pr(\alpha \text{ loses } \& \ n_w \text{ winners}|A)(-\gamma + l_{\hat{\xi}}(n_w))\Big] + \Pr(0 \text{ winners}|A) \cdot \left[-\gamma + \frac{\delta}{K}\left(E_A(\mathbf{p},\hat{\chi},k) + \sum_{\substack{l=1 \\ l\neq k}}^{K} E_{\hat{\chi}(l)}(\mathbf{p},\hat{\chi},l)\right)\right] \tag{4.2}$$

$$E_Y(\mathbf{p},\hat{\chi},k) := \sum_{n_w=1}^{\min(n,c)} \Pr(n_w \text{ winners}|Y) \cdot l_{\hat{\xi}}(n_w) + \Pr(0 \text{ winners}|Y) \cdot \frac{\delta}{K}\left(E_Y(\mathbf{p},\hat{\chi},k) + \sum_{\substack{l=1 \\ l\neq k}}^{K} E_{\hat{\chi}(l)}(\mathbf{p},\hat{\chi},l)\right) \tag{4.3}$$

Lemma 19. *For any strategy $\hat{\chi}$ and signal k, the functions E_A and E_Y are continuous in $\mathbf{p} \in \langle 0,1\rangle^K$.*

Proof. The probabilities $\Pr(n_w \text{ winners}|A)$ and $\Pr(n_w \text{ winners}|Y)$ are continuous. The functions E_A and E_Y are sums of products of continuous functions, so they must be themselves continuous. □

Lemma 20. *Functions E_A and E_Y are well-defined for any $\hat{\chi}$, k and $\mathbf{p} \in \langle 0,1\rangle^K$.*

Proof. For fixed $\mathbf{p}$, $\hat{\chi}$, γ and δ the functions E_A, E_Y define each a system of K linear equations. We can write this system as $\mathbf{A}\mathbf{E}_{\hat{\chi}} = \mathbf{b}$, where $\mathbf{E}_{\hat{\chi}}$ is a vector of corresponding payoff functions $E_{\hat{\chi}(k)}$, and $\mathbf{b} \in \mathbb{R}^K$. The matrix $\mathbf{A}$ is defined as

$$\mathbf{A} := \mathbf{I} - \frac{\delta}{K}(\Pr(0 \text{ winners}|\hat{\chi}(1)), \ldots, \Pr(0 \text{ winners}|\hat{\chi}(K))) \cdot \mathbf{1}^T \tag{4.4}$$

where $\mathbf{I}$ is a $K \times K$ unit matrix and $\mathbf{1}^T$ a K-dimensional row vector of all 1.

This system of equations has a unique solution if the matrix $\mathbf{A}$ is non-singular. This is equivalent to saying that $\det(\mathbf{A}) \neq 0$.

The matrix $\mathbf{A}$ is diagonally dominant, that is $a_{ii} > \sum_{j=1, j\neq i}^{K} |a_{ij}|$. This is because $0 < \delta < 1$, and all the probabilities $\Pr(n_w = c|\hat{\chi}(k)) \leq 1$. It is known that diagonally dominant matrices are non-singular (Taussky (1949)). Therefore, a unique solution $\mathbf{E}_{\hat{\chi}}$ of the system exists and the functions E_A, E_Y are well-defined. □

Suppose that given the probability vector $\mathbf{p}$, there is a deterministic best-response strategy for agent α $\hat{\chi}_{\mathbf{p}}$.

Theorem 21. *If the functions $E_A(\mathbf{p}, \hat{\chi}_{\mathbf{p}}, k)$ and $E_Y(\mathbf{p}, \hat{\chi}_{\mathbf{p}}, k)$ are well-defined and continuous in any p_k, there exists a probability vector $\mathbf{p}^* = (p_1^*, p_2^*, \ldots, p_K^*)$ such that when for signal k, every loser accesses a resource with probability p_k^*, each agent plays a best-response to the strategy of everyone else.*

Proof. Fix γ, δ, $\hat{\chi}$ and $\mathbf{p}$ for all $l \in \{1, \ldots, K\}$, $l \neq k$.

Let $p_k = 0$. If $E_Y \geq E_A$, every loser is best off playing Y and it is a symmetric best-response.

If not, then let $p_k = 1$. If in this case $E_A \geq E_Y$, every loser is best off playing A and again this is a symmetric best-response.

Finally, if both $E_Y < E_A$ for $p_k = 0$, and $E_Y > E_A$ for $p_k = 1$, then from the fact that both functions are well-defined and continuous for $0 \leq p_k \leq 1$, they must intersect for some $0 < p_k^* < 1$. For such p_k^*, the agents are indifferent between actions A and Y. Therefore, it is a symmetric best-response when all the losers play A with probability p_k^*.

We now know that for any coordination signal k, there exists a symmetric best-response given any set strategies $\hat{\chi}(l)$ for other coordination signals $l \neq k$. Therefore, there must exist a probability vector $\mathbf{p}^*$ such that for all coordination signals it is a symmetric best-response to access with these probabilities. □

Intuitively, Theorem 21 shows that for any uniform equilibrium convention $\hat{\xi}^*$, there exists a uniform equilibrium implementation $\hat{\pi}^*$. The overall strategy is then a subgame-perfect equilibrium of the repeated resource allocation game, as shown in Lemma 9.

To illustrate the different equilibrium payoffs agents can get when they adopt different conventions, consider the resource allocation game with $N = 4$ agents and $C = 1$ (to simplify the presentation, assume that $K = 1$). Assume that before round t, no resource has been claimed yet, so there are $n = 4$ losers and $c = 1$ unclaimed resource. If some agent becomes a winner in round t, the agents adopt an extended uniform convention that prescribes their strategies from then on.

For comparison, assume that the agents can adopt either a convention $\hat{\xi}_1$, or a convention $\hat{\xi}_2$. If they adopt convention $\hat{\xi}_1$, the winners have an expected payoff $w_{\hat{\xi}_1} = 4$, and the losers an expected payoff $l_{\hat{\xi}_1} = 0$. On the other hand, if they adopt convention $\hat{\xi}_2$, the winners have an expected payoff $w_{\hat{\xi}_2} = 2$, and the losers an expected payoff $l_{\hat{\xi}_2} = 1$.

Figure 4.2 shows the expected payoff functions (E_A^1 and E_Y^1 for the convention $\hat{\xi}_1$, and E_A^2 and E_Y^2 for the convention $\hat{\xi}_2$), depending on the access probability p. We can see that the equilibrium implementation payoff E_2^* of the convention $\hat{\xi}_2$ is higher than the equilibrium payoff E_1^* of the convention $\hat{\xi}_1$, even though the sum of the winner and loser payoffs is higher

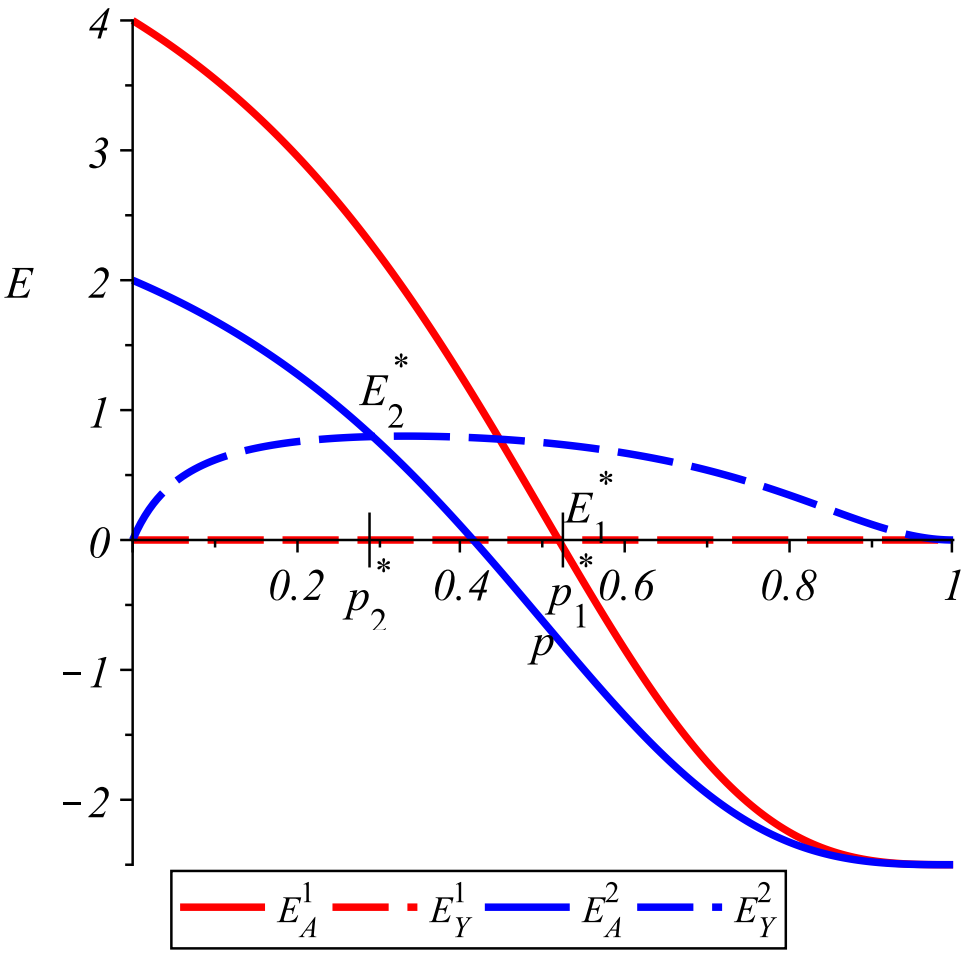


Figure 4.2: Example of expected payoff functions for resource allocation game with $N = 4$ agents, $C = 1$ resources, cost of collision $\gamma = 2$ and discount factor $\delta = 0.8$, given the access probability p. The function E_A^1 and E_Y^1 are expected payoff functions of accessing and yielding, when the agents use an extended convention $\hat{\xi}_1$. Similarly, E_A^2 and E_Y^2 are expected payoff functions when the agents use an extended convention $\hat{\xi}_2$. Convention $\hat{\xi}_1$ has an expected winner payoff $w_{\hat{\xi}_1} = 4$, and expected loser payoff $l_{\hat{\xi}_1} = 0$. Convention $\hat{\xi}_2$ has an expected winner payoff $w_{\hat{\xi}_2} = 2$ and expected loser payoff $l_{\hat{\xi}_2} = 1$.
In the equilibrium implementation π_1 of the convention $\hat{\xi}_1$, the agents access the resource with probability p_1^*, and their expected payoff is $E_1^* = 0$. In the equilibrium implementation π_2 of the convention $\hat{\xi}_2$, the agents access the resource with probability $p_2^* < p_1^*$, and their expected payoff is $E_2^* > E_1^* = 0$.

for convention $\hat{\xi}_1$. This is because the loser receives a positive payoff when the agents adopt a convention $\hat{\xi}_2$; the agents are less likely to "fight" to become a winner, and they access the resource with a lower probability $p_2^* < p_1^*$. This way, there will be less collisions, and the agents will receive a higher expected social payoff when they adopt the convention $\hat{\xi}_2$.

Clearly, any convention where the losers receive a positive payoff will have a higher equilibrium payoff than a convention where the losers' payoff is 0. Intuitively, this is similar to the result of Kuzmics et al. (2010) that in the Nash demand game, a socially efficient convention must be *strong ex-post payoff symmetric* (see Section 2.3). However, it remains an open question whether improving the losers payoff in general improves the equilibrium payoff, since increasing the loser payoff usually leads to a decreasing winner payoff.

4.1.1 Calculating the Equilibrium

While the symmetric subgame perfect equilibrium is guaranteed to exist, in order to actually play it, the agents need to be able to calculate it. It is not always possible to obtain the closed form of the probability of accessing a resource. Therefore, we will show how to calculate the equilibrium strategy numerically.

Let $\mathbf{p}$ be a probability vector, $\hat{\chi}$ a strategy and k a signal. Let $\mathbf{p}_0 := (p_1, p_2, \ldots, p_k = 0, \ldots, p_K)$, i.e. vector $\mathbf{p}$ with p_k set to 0. Let $\mathbf{p}_1 := (p_1, p_2, \ldots, p_k = 1, \ldots, p_K)$. From Theorem 21 we know that either $E_Y(\mathbf{p}_0, \hat{\chi}, k) > E_A(\mathbf{p}_0, \hat{\chi}, k)$, or $E_A(\mathbf{p}_1, \hat{\chi}, k) > E_Y(\mathbf{p}_1, \hat{\chi}, k)$ or the two functions intersect for some $0 \leq p_k \leq 1$. Furthermore, we know that $E_A(\mathbf{p}_0, \hat{\chi}, k) = w_{\hat{\xi}}(c)$ since the probability of successfully claiming a resource is 1 when everyone else yields, and also $E_Y(\mathbf{p}_0, \hat{\chi}, k) = 0$. Therefore, $E_Y(\mathbf{p}_0, \hat{\chi}, k) > E_A(\mathbf{p}_0, \hat{\chi}, k)$ *iff* $w_{\hat{\xi}}(c) > 0$.

W.l.o.g, we will assume that $w_{\hat{\xi}}(c) > 0$. Algorithm 1 shows then how to calculate the probability vector.

Algorithm 1 Calculating the equilibrium probabilities

for Each subset $S \subseteq \{1, 2, \ldots, K\}$ **do**
  Let Σ be a system of equations
  $\forall i \notin S$, Σ contains two equations for $E(\mathbf{p}, \hat{\chi}, i)$. One corresponding to $E_A(\mathbf{p}, \hat{\chi}, i)$, one to $E_Y(\mathbf{p}, \hat{\chi}, i)$.
  $\forall j \in S$, we set $p_j := 1$. Σ contains only one equation for $E(\mathbf{p}, \hat{\chi}, j)$, corresponding to $E_A(\mathbf{p}, \hat{\chi}, j)$.
  So Σ is a system of $2K - |S|$ equations with $2K - |S|$ variables.
  Solve numerically the system of equations Σ.

  if there exists a solution to Σ for which $\forall i \notin S$, $0 \leq p_i \leq 1$ **then**
    We have found a solution
    break;
  end if
end for

4.2 Conventions

In the previous section, we have shown that we can find a symmetric way to reach any equilibrium convention, provided the agents access the resources with a certain probability. We have also shown how to calculate the resource access probability in every stage of the game. In this section, we would like to show specific examples of the conventions that agents can adopt, and discuss their properties.

4.2.1 Bourgeois Convention

The bourgeois convention is the simplest one. Once an agent has accessed a resource successfully for the first time, he will keep accessing it forever. We say that the agent has *claimed* the resource. We don't need any coordination signal to implement it, so we can set $K := 1$.

For N agents and C resources, we will describe the decision problem from the point of view of agent α. Let c be the number of resources that have not been claimed yet, and $n := N - C + c$ the number of agents who have not claimed a resource yet. We define $E(c, \tau_{-\alpha})$ as the expected payoff of the best response strategy for agent α given the strategies $\tau_{-\alpha}$ of all the opponents.

Lemma 22. *For any $\tau_{-\alpha}$ and $\forall c \geq 1$, $E(c, \tau_{-\alpha}) \geq 0$.*

Proof. No matter what is the strategy of the opponents, if agent α chooses to always yield, its payoff will be 0. □

Lemma 23. *If the opponents' strategies $\tau_{-\alpha}$ are such that the agent α is indifferent in every round between yielding and accessing, $E(c, \tau_{-\alpha}) = 0$ for all $c \geq 1$.*

Proof. If the agent α is indifferent between actions Y and A in every round, that means that it is indifferent between a strategy that prescribes Y in every round and any other strategy. The (expected) payoff of the strategy that prescribes always Y is 0. Therefore, the expected payoff of any other strategy must be 0 as well. □

For the purpose of our problem, all the unclaimed resources are identical. Therefore the only parameter of the agent strategy is the probability with which it decides to access – the resource itself is then chosen uniformly at random. Lemma 23 shows a necessary condition for agent α to be indifferent. The following lemma shows a sufficient condition:

Lemma 24. *Assume at round r there are c unclaimed resources. Then there exists a unique $0 \leq p^* \leq c$ such that if all opponents who haven't claimed any resource yet play A with probability $p_c^* = c\left(1 - \sqrt[n-1]{\frac{|\gamma|}{|\gamma| + \frac{1}{1-\delta}}}\right)$, agent α is indifferent between yielding and accessing.*

Proof. From Lemma 23 we know that when agent α is indifferent, it must be that $E(c, \tau_{-\alpha}) = 0$ for all $c \geq 1$.

The expected profit to agent α from playing A and then following best-response strategy (with zero payoff) is

$$E_A(c,\tau_{-\alpha}) = \left(1-\frac{p}{c}\right)^{n-1}\cdot\frac{1}{1-\delta} + \left[1-\left(1-\frac{p}{c}\right)^{n-1}\right]\cdot(-\gamma) \tag{4.5}$$

Here p is the probability with which the opponents access. We want $E_A(c,\tau_{-\alpha}) = E_Y(c,\tau_{-\alpha}) = 0$. This holds if p_c^* is defined as in the theorem above.

Function E_A is decreasing in p on the interval $[0,c]$, while function E_Y is constantly 0. Therefore, the intersection is unique on an interval $[0,c]$. □

Lemma 25. *Assume that all the opponents who haven't claimed any resource access a resource with probability $p < p_c^*$. Then it is best-response for agent α to access.*

Proof. The probability that agent α claims successfully a resource after playing A is

$$\Pr(\text{claim some resource}|A) := \left(1-\frac{p}{c}\right)^{n-1} \tag{4.6}$$

This probability increases as p decreases. Therefore the expected profit of accessing is increasing, whereas the profit of yielding stays 0. □

Theorem 26. *Define an agent's strategy τ as follows: If there are c unclaimed resources, play A with probability $p_c := min\left(1, p_c^*\right)$ (where p_c^* is defined in Lemma 24). Then a joint strategy profile $\tau = (\tau_1, \tau_2, \ldots, \tau_N)$ where $\forall c, \tau_c = \tau$ is a subgame perfect equilibrium of the infinitely repeated resource allocation game.*

Proof. If $p_c^* < 1$, any agent is indifferent between playing Y and playing A, therefore will happily follow strategy τ. If $1 = p_c < p_c^*$, it is best response for any agent to play A, just as the strategy τ prescribes. □

Theorem 27. *For all $c \in \mathbb{N}$, if $p_c = p_c^*$, $E(c,\tau_{-\alpha}) = 0$.*

Proof. We will proceed by induction.

For $c = 0$, the expected payoff is trivially $E(0,\tau_{-\alpha}) = 0$, because there are no free resources.

Let $\forall j < c, E(j,\tau_{-\alpha}) = 0$ and $p_c = p_c^*$. If agent α plays Y, the expected payoff is clearly 0 (it will be 0 now and 0 in the future from the induction hypothesis). If agent α plays A, the expected

payoff is

$$E_A(c,\tau_{-\alpha}) := \left(1-\frac{p_c}{c}\right)^{n-1}\cdot\frac{1}{1-\delta} + \left[1-\left(1-\frac{p_c}{c}\right)^{n-1}\right]\cdot(-\gamma)+\delta\sum_{j=0}^{c}q_{cj}E(j) \tag{4.7}$$

Because of the way the p_c^* is defined, and from the induction hypothesis $E(j,\tau_{-\alpha})=0$ for $j<c$, we get

$$\begin{aligned} E_A(c,\tau_{-\alpha}) &:= \delta q_{cc}E(c,\tau_{-\alpha}) \\ &= \delta q_{cc}\max\{E_A(c,\tau_{-\alpha}),E_Y(c,\tau_{-\alpha})\} \end{aligned} \tag{4.8}$$

Since $\delta q_{cc}<1$, it must be that $E_A(c,\tau_{-\alpha})=0$. □

Theorem 28. *If $p_c<p_c^*$, $E(c,\tau_{-\alpha})>0$.*

Proof. From Lemma 25 we know that when $p_c<p_c^*$, it is a best response to access, so $E(c,\tau_{-\alpha})=E_A(c,\tau_{-\alpha})$. From Lemma 22 we know that for all j, $E(j)\geq 0$. If $p_c<p_c^*$, from the definition of $E_A(c,\tau_{-\alpha})$ (Equation 4.7) we see that $E(c,\tau_{-\alpha})>0$. □

Theorem 28 shows that if we have enough resources so that $p_c^*\geq 1$, the expected payoff for the agents, even when they access all the time, will be positive.

Given the number of agents N, discount factor δ and collision cost γ, the necessary number of resources c^* for the expected to be positive is:

$$c^* := \frac{1}{1-\sqrt[n-1]{\frac{|\gamma|}{|\gamma|+\frac{1}{1-\delta}}}} \tag{4.9}$$

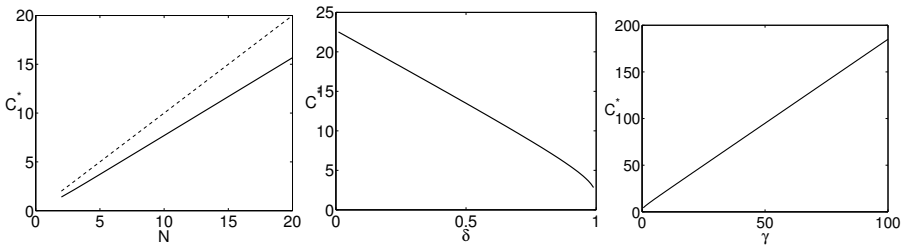
Figure 4.3 illustrates the value of c^* depending on N, δ, and γ respectively.

Let us now look at the price of anonymity for the bourgeois convention (as defined in Definition 24). The highest social payoff any strategy profile τ can achieve in an N-agent, C-resource allocation game ($N\geq C$) is

$$\max E(\tau) := \frac{C}{1-\delta}. \tag{4.10}$$

This is achieved when in every round, every resource is accessed by exactly one agent. Such strategy profile is obviously asymmetric.

If each agent knew which part of the bourgeois convention to play at the beginning of the game, this convention would be socially efficient. However, when the agents are anonymous, they have to learn which part of the convention they should play through randomization. For



(a) N (b) δ (c) γ

Figure 4.3: Minimum number of resources c^* needed for the expected payoff of bourgeois convention to be positive, depending N, δ, and γ. One parameter is varying, the other parameters are set to $N = 10$, $\delta = 0.8$, $\gamma = 2$. For varying N, the dashed line shows when $c = N$.

the bourgeois convention for small C, this randomization wipes out all the efficiency gains. Therefore, its price of anonymity is infinite.

4.2.2 Market Convention

We saw that the bourgeois convention leads to zero expected social payoff for a small number of resources. We would like to improve the expected payoff here. In the bourgeois convention, the agents receive zero expected payoff because the demand for resources is too high compared to the supply. We need to decrease the demand, while increasing the supply. This is often achieved through markets. Shneidman et al. (2005) present some of the reasons why markets might be appropriate for resource allocation.

We assume the following:

- Agents can observe $K \geq 1$ coordination signals.
- Agents have a decreasing marginal utility when they access a resource more often. More precisely, successfully accessing some resource for each additional signal will have less additional utility compared to the previous signals (see Figure 4.4 for an illustration).
- They pay a fixed price per each successful access, to the point that each agent prefers to access a resource only for one signal out of K. In practice, this could be implemented by a central authority that observes the convergence rate of the agents, and dynamically increases or decreases the price to achieve convergence.

Such assumptions define what we call “market” convention, where the winners only access their claimed resource for the signals they observed when they first claimed it. The price the agents have to pay serves to decrease the demand. The coordination signal effectively increases the supply of resources K-times, because the resource allocation may be different for each of the signal values.

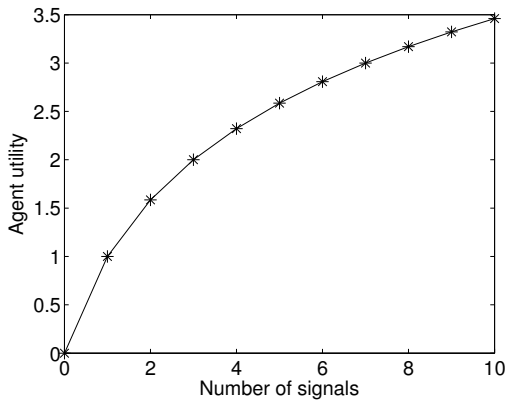


Figure 4.4: An example of a decreasing marginal utility function, given the number of signals for which an agent successfully accesses some resource.

We know that we can implement this convention for $C \geq 1$ resources using symmetric play (see Section 4.1). We can also use Algorithm 1 to calculate the access probabilities. For the ease of exposition, we will first describe the market convention for $C = 1$ resource. Then we will generalize the description to $C > 1$ resources.

One Resource

When each agent only accesses the resource for one signal, we need $K = N$ signals to make sure everyone gets to access once.

In the N-agent, 1-resource case, imagine there are still n agents playing and $(N-n)$ agents who have already claimed the resource for some signal. Imagine that the n agents observe one of the n signals for which no resource has been claimed.

Assume that all agents access the resource with probability p_n. The expected payoff of accessing a resource for agent α is

$$\begin{aligned} E_A(p_n, n) := (1-p_n)^{n-1} \cdot \left(1 + \frac{\delta}{N} \cdot \frac{1}{1-\delta}\right) \\ + \left[1 - (1-p_n)^{n-1}\right] \cdot \left[-\gamma + \frac{\delta n}{N - \delta(N-n)} E_A(p_n, n)\right] \end{aligned} \tag{4.11}$$

The expected payoff of yielding for agent α is

$$\begin{aligned} E_Y(p_n, n) := (n-1)\, p_n (1-p_n)^{n-2} E(n-1) \\ + \left[1 - (n-1)\, p_n (1-p_n)^{n-2}\right] \frac{\delta n}{N - \delta(N-n)} E_Y(p_n, n) \end{aligned} \tag{4.12}$$

When $p_n = 1$, accessing a resource will always lead to a collision, so the payoff of accessing will

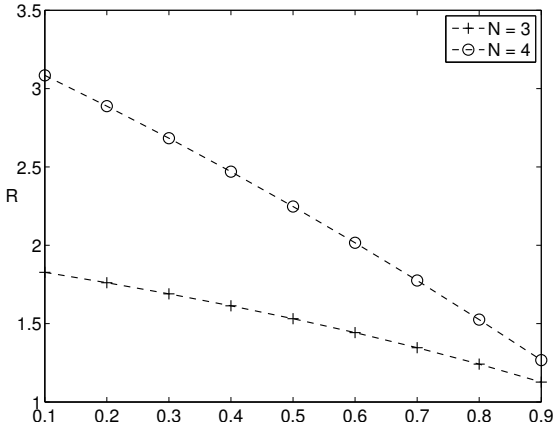


Figure 4.5: Market convention: Price of anonymity for $C = 1$, $K = N$, $\gamma = 0.5$ and varying δ.

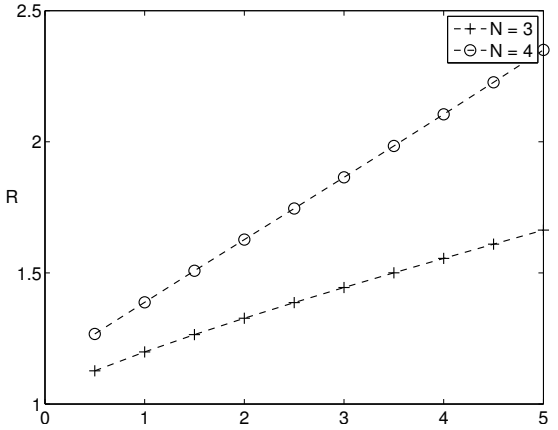


Figure 4.6: Market convention: Price of anonymity for $C = 1$, $K = N$, $\delta = 0.9$ and varying γ.

be negative. When $p_n = 0$, accessing a resource will always claim it, so the payoff of accessing will be positive. In the equilibrium, the agents should be indifferent between accessing and yielding. Therefore, we want to find p_n^* such that $E_A(p_n^*, n) = E_Y(p_n^*, n) = E(n)$.

Finding a closed form expression for p_n^* is difficult, but we can use Algorithm 1 to calculate this probability, as well as the expected payoff $E(n)$, numerically.

Figures 4.5 and 4.6 show the price of anonymity of the market convention (as defined in Definition 24) of the market convention for varying discount factor δ, and varying cost of collision γ, respectively. From Section 4.2.1, we saw that the price of anonymity for $C = 1$ is ∞. On the contrary, for the market convention this price is in both cases finite and relatively small.

Multiple Resources

Assume now that $C \geq 1$. In any given round, we will denote $\mathbf{c} := (c_1, c_2, \ldots, c_K)$ the vector of resources which have not been claimed yet for each value of the coordination signal $k \in \{1, 2, \ldots, K\}$.

For a given vector of unclaimed resources c and a probability vector $\mathbf{p} := (p_1, p_2, \ldots, p_K)$, we will define the expected payoff functions $E_A(n, \mathbf{c}, \mathbf{p}, k)$ and $E_Y(n, \mathbf{c}, \mathbf{p}, k)$ of agent α if she takes actions A and Y, respectively. Here n is the number of agents who have not claimed any resource yet, and k is the signal the agents observe in the current round.

From Theorem 21 we know that there exists an equilibrium for the market convention. For vector of unclaimed resources $\mathbf{c}$ we can define two equilibrium payoff functions: $E(n, \mathbf{c}, k)$ the expected payoff when the agents observe signal k, and $E(n, \mathbf{c})$ the expected payoff *before* the agents observe the coordination signal. We can see that

$$E(n, \mathbf{c}) := \frac{1}{K} \sum_{k=1}^{K} E(n, \mathbf{c}, k).$$

The functions $E_A(n, \mathbf{c}, \mathbf{p}, k)$ and $E_Y(n, \mathbf{c}, \mathbf{p}, k)$ are defined as follows:

$$\begin{aligned} E_A(n, \mathbf{c}, \mathbf{p}, \chi, k) := {} & \Pr(\alpha \text{ wins} | A) \cdot w + (1 - \Pr(\alpha \text{ wins} | A)) \cdot (-\gamma) \\ & + \sum_{n_w=1}^{\min(c_k, n)} \Pr(\alpha \text{ loses, } n_w \text{ winners} | A) \cdot \delta \cdot E(n - n_w, (c_1, \ldots, c_k - n_w, \ldots, c_K)) \\ & + \Pr(\alpha \text{ loses, } n_w = 0 | A) \cdot \left[\frac{\delta}{K} \cdot \sum_{l=1}^{K} E_{\chi_l}(n, \mathbf{c}, \mathbf{p}, l, \chi) \right] \end{aligned} \tag{4.13}$$

$$\begin{aligned} E_Y(n, \mathbf{c}, \mathbf{p}, \chi, k) := {} & \sum_{n_w=1}^{\min(c_k, n)} \Pr(n_w \text{ winners} | Y) \cdot \delta \cdot E(n - n_w, (c_1, \ldots, c_k - n_w, \ldots, c_K)) \\ & + \Pr(n_w = 0 | Y) \cdot \frac{\delta}{K} \cdot \sum_{l=1}^{K} E_{\chi_l}(n, \mathbf{c}, \mathbf{p}, l, \chi) \end{aligned} \tag{4.14}$$

Here the winner payoff w is defined as $w := 1 + \frac{\delta}{K \cdot (1-\delta)}$. This is because the winner will transmit for only one signal: once in the current round, and than in any future round with probability $\frac{1}{K}$.

What are the probabilities that there will be n_w winners in each of the cases? We will start with the simplest case, $\Pr(n_w \text{ winners} | Y)$, given that there are n agents (including agent α), c_k resources and all agents except α play action A with probability p_k.

The problem of calculating this probability is very similar to the well-known *balls-and-bins*

problem (Raab and Steger (1998)). In the balls-and-bins problem we assume that we have n balls who are each randomly assigned into one of the c bins. The goal is to find a probability that i bins will have exactly one ball in them. We will express this probability as $\phi(n,c,i)$.

There are N_i ways to pick some i balls and place them into some i bins so that each bin has one ball,

$$N_i := \binom{c}{i}\binom{n}{i} \cdot i! \cdot (c-i)^{n-i} \tag{4.15}$$

The total number of ways to place n balls in c bins so that exactly i have one ball can be then obtained from the *generalized inclusion-exclusion* principle:

$$\begin{aligned}\sum_{j=i}^{\min(c,n)} (-1)^{j-i}\binom{j}{i}N_j &= \sum_{j=i}^{\min(c,n)} (-1)^{j-i}\binom{j}{i}\binom{c}{j}\binom{n}{j} \cdot j! \cdot (c-j)^{n-j} \\ &= n!\binom{c}{i}\sum_{j=i}^{\min(c,n)} (-1)^{j-i}\binom{c-i}{j-i}\frac{(c-j)^{n-j}}{(n-j)!}\end{aligned} \tag{4.16}$$

In the simplification above, we use the *absorption identity* $\binom{j}{i}\binom{c}{j} = \binom{c}{i}\binom{c-i}{j-i}$.

There are a total of c^n ways to arrange n balls into c bins. Therefore, the probability $\phi(n,c,i)$ is

$$\phi(n,c,i) := \frac{n!}{c^n}\binom{c}{i}\sum_{j=i}^{\min(c,n)} (-1)^{j-i}\binom{c-i}{j-i}\frac{(c-j)^{n-j}}{(n-j)!} \tag{4.17}$$

How can we use the function ϕ to calculate $\Pr(n_w \text{ winners}|Y)$? The $n-1$ agents (other than α) decide to play action A with probability p_k, and then choose the resource to access randomly. The agents who choose to access a resource correspond to the balls-and-bins problem. Therefore,

$$\Pr(n_w \text{ winners}|Y) := \sum_{i=0}^{n-1}\binom{n-1}{i} p_k^i \cdot (1-p_k)^{n-1-i} \cdot \phi(i,c,n_w). \tag{4.18}$$

To calculate the probability $\Pr(\alpha \text{ wins}|A)$, we can proceed as follows. We assume w.l.o.g that α accesses resource 1. There will be some i agents (out of $n-1$) who will choose action A. We then need all of them to choose other resource than 1. Therefore,

$$\Pr(\alpha \text{ wins}|A) := \sum_{i=0}^{n-1}\binom{n-1}{i} \cdot p_k^i \cdot (1-p_k)^{n-1-i}\left(1-\frac{1}{c}\right)^i \tag{4.19}$$

Finally, to calculate the probability $\Pr(\alpha \text{ loses}, n_w \text{ winners}|A)$, we can use again the balls-and-bins problem. Given that there are $0 \le i \le n-1$ agents who choose action A, there will be

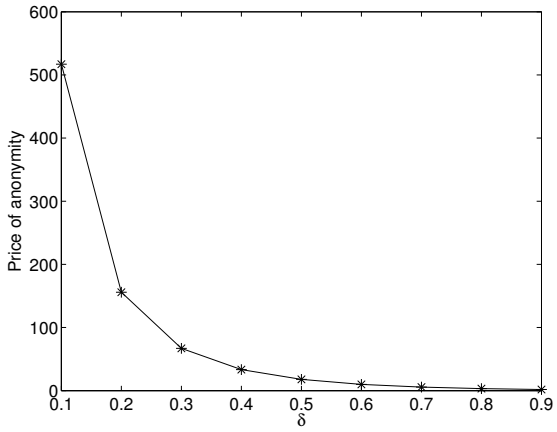


Figure 4.7: Market convention: Price of anonymity for $N = 6$, $C = 3$, $K = 2$, $\gamma = 0.5$ and varying δ.

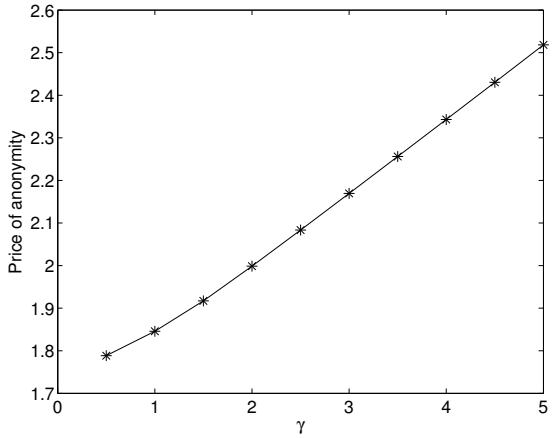


Figure 4.8: Market convention: Price of anonymity for $N = 6$, $C = 3$, $K = 2$, $\delta = 0.9$ and varying γ.

$0 \leq j \leq i$ agents who choose the same resource as agent α. The remaining $(i-j)$ agents face the same balls-and-bins problem for $c-1$ bins (1 bin is already occupied by agent α). Therefore,

$$\Pr(\alpha \text{ loses, } n_w \text{ winners}|A) := \sum_{i=1}^{n-1} \binom{n-1}{i} p_k^i (1-p_k)^{n-1-i} \sum_{j=1}^{i} \binom{i}{j} \left(\frac{1}{c}\right)^j \left(1-\frac{1}{c}\right)^{i-j} \cdot \phi(i-j, c-1, n_w) \tag{4.20}$$

Now that we have expressed the expected payoff functions E_A and E_Y explicitly, we can use Algorithm 1 to calculate the equilibrium access probabilities and expected payoffs.

Figures 4.7 and 4.8 show the price of anonymity of the market convention for $C = 3$, $K = 2$ and $N = C \cdot K = 6$. When the discount factor δ grows, the price of anonymity decreases (note that in Figure 4.7 the y-axis is logarithmic). This is because for small δ, the benefit of winning

the resource right away is much higher than the payoff of winning later. On the other hand, as δ gets closer to 1, the agents don't care whether they win now or later. Since the market convention guarantees that everyone will be able to access some resource for some signal value, when $\delta \to 1$, the expected payoff of winner and losers will be the same. Also, as $\delta \to 1$, the cost the agents have to pay for learning the convention decreases compared to the payoff they obtain after they have learnt it.

When γ increases, the price of anonymity increases. The cost of collision has a direct effect on the expected payoff functions E_A and E_Y. Therefore, the expected equilibrium payoff will be higher if the cost is lower. Changing the γ has no effect on the optimal asymmetric outcome though, since the agents don't have to pay any cost because there are no collisions.

Calculating the equilibrium access probabilities for the market convention is difficult – we need to use a numerical algorithm, and as the number of signals K grows, the number of equations grows exponentially. So we would like to find access probabilities which are easy to compute and for which the agents' incentive to deviate is *too small*. Indeed, game theory is often interested in ϵ*-equilibria*, in which no agent can improve her payoff by more than $\epsilon > 0$.

The market pricing ensures that each agent only wants to access a resource for one signal value. It also doesn't depend on the access probabilities of the agents, only on their utility functions. Once the agents converge to the asynchrony round (i.e. a pure-strategy NE of the resource allocation for every signal value), their future expected payoff will be

$$\frac{K^{-1}}{1-\delta}, \tag{4.21}$$

and no agent can improve her payoff by deviating since the players are playing a PSNE of the stage game in each round.

If the agents who haven't claimed their resource yet play action A with a constant probability $0 < p_{const} < 1$, the expected time before the reach the asynchrony is finite. (from the properties of the *balls-and-bins* problem, see Section 4.2.3, or Raab and Steger (1998)). We can prove the following theorem:

Theorem 29. *Suppose that in the N-agent, C-resource allocation game, the agents adopt the market convention with the following implementation: The agents who haven't claimed any resource yet play action A with a constant probability p_{const} (we call this the* constant-probability *implementation). Let $E(\delta)$ be the expected payoff for each agent in this case for a given discount factor δ. Let $E'(\delta)$ be the expected payoff of the best-response strategy to this convention and implementation.*

Then for any $\epsilon > 0$, there exists $0 < \delta_0 < 1$ such that for all δ, $\delta_0 \leq \delta < 1$,

$$\frac{E(\delta)}{E'(\delta)} > 1 - \epsilon. \tag{4.22}$$

Proof. Because of the market pricing, each agent only wants to access one resource for one value of the coordination signal. So the best-response payoff E' is

$$E'(\delta) \leq \frac{K^{-1}}{1-\delta}, \tag{4.23}$$

no matter what strategy do the other agents play.

When the agents adopt the market convention with the constant-probability implementation, then in every round until they converge to a PSNE, they receive a payoff between $\gamma < 0$ (the collision cost) and 1. After they reach the PSNE, their expected payoff is

$$\frac{K^{-1}}{1-\delta} \tag{4.24}$$

as stated above. We can therefore say that

$$E(\delta) \geq \sum_{i=0}^{\infty} \Pr(\text{agents reach PSNE in } i \text{ steps}) \cdot \left[\gamma \cdot \frac{1-\delta^i}{1-\delta} + K^{-1} \cdot \frac{\delta^i}{1-\delta}\right] \tag{4.25}$$

We can define a random variable X such that $X = i$ if the agents reach a PSNE after exactly i steps. From the properties of the expected value, we can ee that

$$E(\delta) \geq \frac{\gamma \cdot (1 - E\left[\delta^X\right]) + K^{-1} \cdot E\left[\delta^X\right]}{1-\delta}. \tag{4.26}$$

The function $\phi(x) := \delta^x$ is a convex function. From the Jensen's inequality (Jensen (1906)), we know that

$$E\left[\delta^X\right] \geq \delta^{E[X]}. \tag{4.27}$$

Therefore,

$$\frac{E(\delta)}{E'(\delta)} \geq \frac{\gamma \cdot \left(1 - \delta^{E[X]}\right) + K^{-1} \cdot \delta^{E[X]}}{K^{-1}}. \tag{4.28}$$

The expected time $E[X]$ to reach the PSNE is finite and doesn't depend on δ, so we can treat is as a constant. Because $\delta^{E[X]}$ is continuous in δ, monotonous and $\lim_{\delta \to 1_-} \delta^{E[X]} = 1$, we can see that for a given $\epsilon > 0$, there exists $0 < \delta_0 < 1$ such that for all δ, $\delta_0 \leq \delta < 1$,

$$\frac{E(\delta)}{E'(\delta)} > 1 - \epsilon. \tag{4.29}$$

□

By ensuring that each agent only wants to access some resource for one signal value, the market convention makes the cooperative strategy from Chapter 3 *almost* rational.

4.2.3 Expected Convergence

In this section, we will analyze what is the expected number of rounds the agents need to converge to a perfect allocation of resources (one where all resources are used by exactly one, and there are no collisions). We will first prove an upper bound on the expected number of steps to the convergence for the bourgeois convention, and then present experiments for the market convention.

Bourgeois Convention

In order to prove the convergence of the bourgeois convention, we will describe its execution as a Markov chain (as we did in Chapter 3).

A Markov chain describing the execution of the bourgeois convention in N-agent, C-resource allocation game is a chain whose state at round t is $X_t \in \{0,1,\ldots,C\}$, where $X_t = c$ means that there are c unclaimed resources at round t. We will again use the Theorem 7 to derive an upper bound on the hitting time. We first need to calculate the expected state $E(X_{t+1}|X_t = c)$.

Lemma 30. *Let $X_t = c$, and let there be $n := N - C + c$ agents who have not claimed a resource yet. Let us denote $q(n,c) = \frac{p}{c} \cdot n \cdot \left(1 - \frac{p}{c}\right)^{n-1}$ that a resource i will be claimed in round t if the agents play the subgame-perfect equilibrium strategy vector described above.*

Then the next expected state is

$$E(X_{t+1}|X_t = c) := \big(1 - q(n,c)\big) \cdot c$$

Proof. For a resource i, we can denote W_i the random variable, where $W_i = 1$ if the resource i has been claimed in round t, and $W_i = 0$ otherwise. The random variable W_i is Bernoulli-distributed with probability $q(n,c)$.

The next expected state is then

$$E(X_{t+1}|X_t = c) = c - E\left[\sum_{i=1}^{c} W_i\right] = c - \sum_{i=1}^{c} E[W_i] = \big(1 - q(n,c)\big) \cdot c, \tag{4.30}$$

because $E[W_i] = q(n,c)$. □

In the following lemmas, we will denote

$$\lambda := \frac{|\gamma|}{|\gamma| + \frac{1}{1-\delta}} \tag{4.31}$$

Lemma 31. *For a given collision cost γ and discount factor δ, there exists a constant $0 < \mu < 1$ such that for $c \leq \mu \cdot n$, $p^* < 1$.*

Proof. According to the definition of the subgame perfect equilibrium strategy, $p^* := c \cdot \left(1 - \sqrt[n-1]{\lambda}\right)$.

We want $p^* < 1$, which is equivalent to

$$c \cdot \left(1 - \sqrt[n-1]{\lambda}\right) < 1 \tag{4.32}$$

$$\left(1 - \frac{1}{c}\right)^{n-1} < \lambda \tag{4.33}$$

$$\tag{4.34}$$

We know that $c \leq \mu \cdot n$, so

$$\left(1 - \frac{1}{c}\right)^{n-1} \leq \left(1 - \frac{1}{\mu \cdot n}\right)^{n-1} \leq e^{-\mu}. \tag{4.35}$$

If we therefore set μ such that $e^{-\mu} < \lambda$, the access probability $p^* < 1$. □

Lemma 32. *For given γ and δ, there exists $0 < \eta < 1$ such that for any c,*

$$E(X_{t+1} | X_t = c) \leq \left(1 - \eta\right) \cdot c$$

Proof. We will prove the lemma for two cases: when $p^* < 1$ and when $p^* = 1$.

First, let us prove the case $p^* < 1$, that is $p^* = c \cdot \left(1 - \sqrt[n-1]{\lambda}\right)$. Therefore, $q(n,c) = \left(1 - \sqrt[n-1]{\lambda}\right) \cdot n \cdot \lambda$. It can be shown that for any n,

$$q(n,c) = \left(1 - \sqrt[n-1]{\lambda}\right) \cdot n \cdot \lambda \geq -\lambda \log \lambda. \tag{4.36}$$

Now let $p^* = 1$. From Lemma 31 it must be that $c > \mu \cdot n$. Then

$$q(n,c) := \frac{c}{n} \cdot \left(1 - \frac{1}{c}\right)^{n-1} \geq \mu \cdot \left(1 - \frac{1}{\mu \cdot n}\right)^{n-1}, \tag{4.37}$$

because $q(n,c)$ is increasing with c.

Now

$$\mu \cdot \left(1 - \frac{1}{\mu \cdot n}\right)^{n-1} \geq \mu \cdot e^{-\mu}. \tag{4.38}$$

For fixed γ, δ, the μ and λ are constants, so we can set η as

$$\eta := \min(\mu \cdot e^{-\mu}, -\lambda \log \lambda). \tag{4.39}$$

From above, this proves the lemma. □

Theorem 33. *The expected time for the agents to converge to a resource allocation where all the resources are claimed is* $O(\log C)$.

Proof. We have shown how we can express the expected convergence time as expected hitting time of a certain Markov chain.

From Lemma 32 we saw that there exists η such that for any c,

$$E(X_{t+1}|X_t = c) \leq \big(1-\eta\big)\cdot c.$$

We can now combine this result with Theorem 7 to show that the expected hitting time from the state C to state 0 is

$$k_N^0 < \lceil \log_{\frac{1}{1-\eta}} C \rceil + \frac{1}{\eta} \approx O\left(\frac{1}{\eta}\cdot \log C\right) = O(\log C), \tag{4.40}$$

because η is a constant.

□

Market Convention

For the market convention, it is unfortunately very difficult to express the expected number of convergence steps in a closed-form expression. However, we can use Theorem 5 to calculate the expected number of convergence steps for a given parameters N, C, K, γ and δ .

The Markov chain for the market convention for K signals and C resources looks as follows: Its state at time t is $V_t \in \{0,1,\ldots,C\}^K$, where V_{t_k} denotes how many resources have not been claimed for signal k. The initial state V_0 is such that $V_{0_k} = C$ for all $k \in \{1,\ldots,K\}$. If $N \geq C\cdot K$, the final state is when $V_{t_k} = 0$ for all k. When $N < C\cdot K$, the final states are such that

The transition probabilities between two states V_i and V_j, $V_i \neq V_j$, are the following: Suppose $\exists k : V_{j_k} < V_{i_k}$ and $\forall l \neq k : V_{j_l} \neq V_{i_l}$. Let us denote $c := V_{i_k}$, i.e. the number of unclaimed resources in state V_i for signal k, and $n := N - (C - V_{i_k})$ the number of agents who have not claimed any resource for signal k in state V_i.

$$\Pr(V_{t+1} = V_j | V_t = V_i) := \frac{1}{K}\sum_{m=0}^{n}\binom{n}{m} p_k^m (1-p_k)^{n-m}\cdot \phi(m,c,V_{i_k} - V_{j_k}) \tag{4.41}$$

Otherwise if $V_j \neq V_i$, $\Pr(V_{t+1} = V_j | V_t = V_i) := 0$.

Figure 4.9 shows the expected number of rounds to converge for varying δ. The influence of δ on the convergence time is negligible. Figure 4.10 shows the convergence for varying collision

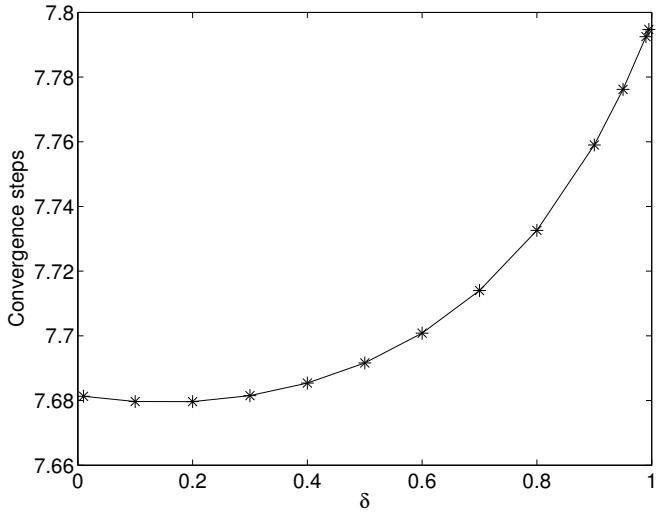


Figure 4.9: Market convention: Expected number of convergence steps given $N = 6$, $C = 3$, $K = 2$, $\gamma = 1.0$ and varying δ.

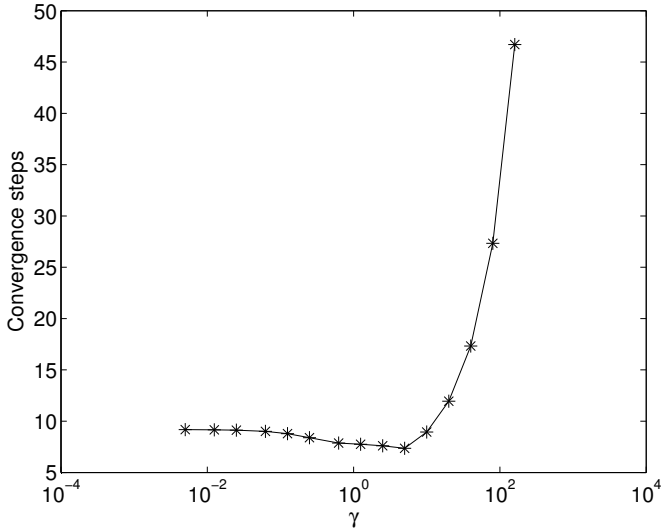


Figure 4.10: Market convention: Expected number of convergence steps given $N = 6$, $C = 3$, $K = 2$, $\delta = 0.9$ and varying γ.

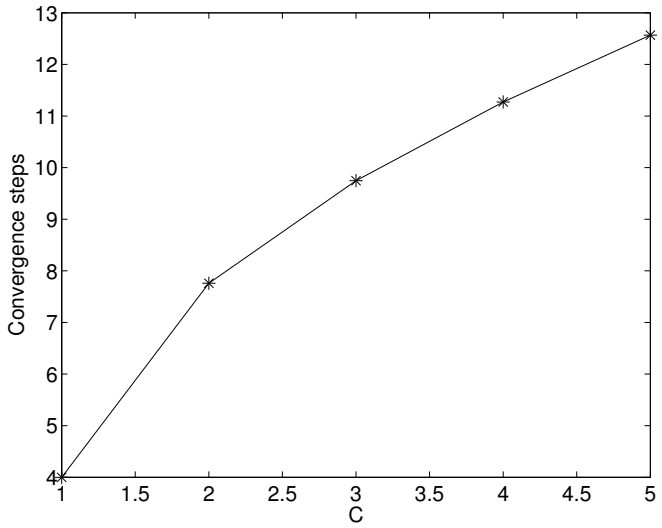


Figure 4.11: Market convention: Expected number of convergence steps given $K = 2$, $\delta = 0.9$, $\gamma = 1.0$ and varying number of resources C and agents $N = 2 \cdot C$.

	ex-post fair	efficient	rational
C&F'11	(✓)[1]	✓	*no*
Bourgeois	*no*	*no*	✓
Egalitarian[2]	✓	✓	✓
Market	✓	?	✓

Table 4.1: Properties of conventions

cost γ. For γ close to 0, the convergence time remains stable. However, for very high cost γ, the convergence time increases linearly with γ. In this case, the high cost of collision drives the resource access probability low, because agents try to avoid collisions "at all costs".

Figure 4.11 shows the expected convergence when we increase number of resources C and number of agents N proportionally. The increase in convergence time is still sub-linear to the increase in C.

4.2.4 Convention Properties

We compare the properties of the following conventions: *C&F'11*, a channel allocation algorithm presented in Chapter 3; *bourgeois*, presented in Section 2.3; *egalitarian*, presented in Section 2.3; and *market*, presented in this chapter.

We compare the conventions according to the following properties:

[1] Fair asymptotically, as $N \to \infty$.
[2] Only for 2-agents, 1-resource games.

Ex-post fairness Is the expected payoff to all agents the same *even after asynchrony*?

Efficiency Does the convention maximize social welfare among all possible conventions?

Rationality Is it an equilibrium for the agents to adopt the convention?

Table 4.1 summarizes the properties of the conventions. The *C&F'11* convention is only approximately ex-post fair. The fairness is improving as the number of coordination signals increases, but some agents might have a worse payoff than others. On the other hand, it is efficient, at least with no discounting ($\delta = 1$). However, it is not rational. The bourgeois convention is neither fair nor efficient, in fact the expected payoff to the agents is 0 (for a small number of resources). It is rational though, since the agents are indifferent between being a winner and a loser. The egalitarian convention is fair, efficient and rational. However, it only works for games of 2 agents and 1 resource. Finally, the market convention is fair and rational. It is clearly more efficient than the bourgeois convention. Nevertheless, finding the most efficient convention remains an open problem.

4.3 Folk theorems and symmetric equilibria

In the previous sections, we have analyzed a special kind of symmetric equilibria of the resource allocation game. The agents first followed a Markovian implementation, and as soon as they play a pure-strategy NE, they adopted a convention. For the general infinitely repeated games with discounting, the so-called *folk theorem* (Theorem 3, Section 2.1.1) characterizes the entire set of payoffs that can be achieved in a Nash equilibrium of the repeated game.

According to the Folk theorem, any convex combination of payoffs achieved in the stage game can be achieved as an average payoff in the infinitely repeated game, provided that the discount factor is high enough and provided the payoffs Pareto-dominate the minimax payoff. For the resource allocation game, the minimax payoff is $(0,0,\ldots,0)$ and is achieved in the mixed strategy Nash equilibrium.

Our focus so far has been on finding symmetric equilibrium strategies. The folk theorem doesn't say anything about whether the equilibrium strategy will be symmetric, even if the payoff vector is symmetric. Nevertheless, we can define another class of symmetric strategies of the infinitely repeated game, than the one based on conventions and their implementations. The strategies have the following form: The players follow a symmetric (mixed) strategy of the stage game. If one player deviates from this strategy, other players punish her by following the minimax strategy. From the Folk theorem, such strategy can be sustained as the Nash equilibrium of the repeated game (though not necessarily a subgame-perfect equilibrium).

A symmetric strategy of the stage resource allocation game is a vector of access probabilities $\mathbf{q} = (q_1, q_2, \ldots, q_C)$ where q_c is the probability that each agent will access resource c. We are interested in finding access probability vector $\mathbf{q}^*$ which achieves the highest expected payoff.

For a given access probability vector $\mathbf{q}$, the expected payoff that an agent receives is as follows:

$$E(\mathbf{q}) := \sum_{j=1}^{C} q_j \cdot \left[(1-q_j)^{N-1} \cdot 1 - \left(1-(1-q_j)^{N-1}\right) \cdot |\gamma|\right] \tag{4.42}$$

Theorem 34. *For a resource allocation game with $N = 2$ agents and $C = 1$ resource, the resource access probability which maximizes the expected payoff of the stage game is*

$$q^* = \frac{1}{2} \cdot \frac{1}{1+|\gamma|}. \tag{4.43}$$

The highest expected payoff of a symmetric strategy in the stage game is then

$$E^* = \frac{1}{4} \cdot \frac{1}{1+|\gamma|}. \tag{4.44}$$

The price of anonymity of the strategy which accesses with probability q^* is then $4 \cdot (1+|\gamma|)$.

Proof. We calculate the derivative of expected payoff function from Equation 4.42 for $N = 2$ and $C = 1$:

$$\frac{\partial E(q)}{\partial q} = 1 - 2q \cdot (1+|\gamma|)$$

Setting the first derivative equal to 0, we get

$$q^* = \frac{1}{2} \cdot \frac{1}{1+|\gamma|}.$$

Since the second derivative is

$$\frac{\partial^2 E(q)}{\partial^2 q} = -1 - |\gamma| < 0,$$

the probability q^* is a point where the expected payoff function $E(q)$ reaches a local maximum. □

For the general case of resource allocation game with N agents and C resources, we can find the probability vector which maximizes the Equation 4.42 (given the constraint $\sum_{j=1}^{C} q_j \leq 1$) using a numerical algorithm.

Figure 4.12 compares the price of anonymity of the folk-theorem-based symmetric strategy with the price of anonymity of the market convention, for $N = 3$ agents and $C = 1$ resource. Since the price of anonymity of the folk theorem strategy doesn't depend on the discount factor δ (it only needs to be high enough for the strategy to be an equilibrium), we only show the graph for varying collision cost γ. The price of anonymity of the folk-theorem strategy is

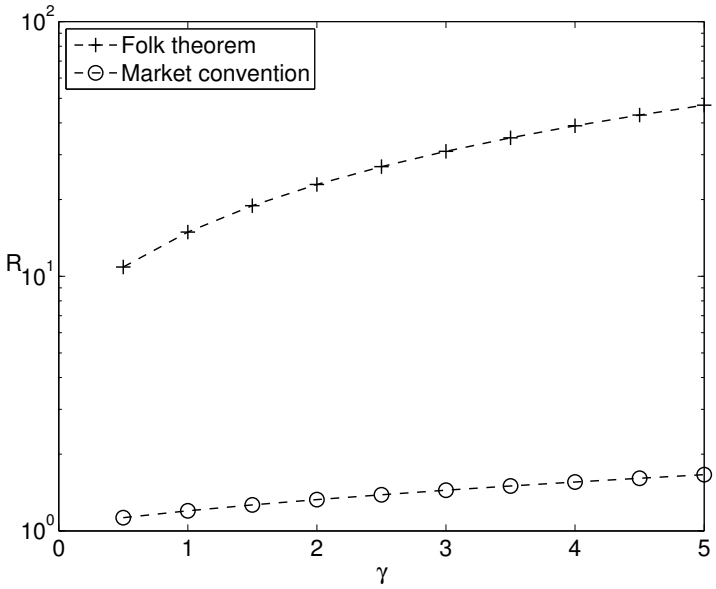


Figure 4.12: Price of anonymity of the symmetric strategy following from the folk theorem, compared to the price of anonymity of the market convention for $N = 3$, $C = 1$ and varying cost of collision γ.

an order of magnitude higher than the price of anonymity of the market convention. We can see that allowing the agents to learn to alternate using a market convention can bring higher social payoff than when they play the most efficient symmetric strategy of the stage game.

4.4 Conclusions

In this chapter we have analyzed symmetric subgame-perfect equilibria of the infinitely repeated resource allocation game where agents discount future payoffs with a discount factor $0 < \delta < 1$. We focused on equilibria with the following general structure (described already in Chapter 2): The agents first start by choosing their actions randomly, and as soon as they reach an efficient resource allocation, they adopt a *convention* that prescribes their play from then on. We have defined the *augmented* convention, that allows the agents to condition their strategy on the value of global coordination signal (such as the one described in Chapter 3). We have shown that for any augmented equilibrium convention of the resource allocation game, there exists an equilibrium implementation. An implementation prescribes the same randomized strategy to all the agents who have not yet successfully accessed some resource. We have presented a numerical algorithm that one can use to calculate the equilibrium implementation strategies for a given convention.

We have defined two conventions, the bourgeois and the market convention. We have analytically shown the equilibrium implementation access probabilities. When the number of resources is low compared to the number of agents, the bourgeois convention leads to zero

equilibrium payoff to the agents. Therefore, its price of anonymity (as defined in Section 2.4) is infinite. The market convention decreases the conflict between the agents and increases the expected payoff. It increases the resource supply by assuming a global coordination signal, that allows the agents to adopt a different resource allocation for each signal value. At the same time, the market convention decreases the demand for resources by charging the agents a price for each successful access of a resource. This way, if we assume that the agents have decreasing marginal utility from accessing more often, we can set the price such that each agent only wants to access a resource for one signal value. For K signals, this increases the capacity K-times. We have proven analytically that for the bourgeois convention, the agents converge to an efficient resource allocation in time logarithmic in the number of resources C.

In the future work, we would like to investigate whether there exist more efficient conventions than the market convention (i.e. conventions with smaller price of anonymity). In general, finding an optimal convention is an NP-hard problem (Balan et al. (2011)), but for a more restricted set of infinitely repeated resource allocation games, we might be able to find the optimal convention, similar to the Thue-Morse sequence (Richman (2001)) used by Kuzmics et al. (2010) in the Nash demand game.

5 Conclusions

In this thesis, we have analyzed problems where multiple independent agents try to access a set of resources, and where each resource can be only used by one agent at a time. Examples of such problems include: wireless networks, where only one device may transmit on a given channel; allocating parking lots, where each parking lot can only be occupied by one vehicle at a time; and auctions, where bidders prefer to participate in an auction with less competition, so as to obtain a lower price for the desired good.

The main difficulty in this kind of resource allocation problems is that all the agents prefer the same outcome – each agent prefers to be the one who can access the resource. But when all the agents try to access the resource simultaneously, they collide. In order to resolve this conflict, the agents need to coordinate their access. This coordination can be either centralized (where a central authority decides on the allocation, and communicates it to the agents), or decentralized, where the agents decide for themselves which resource they should access, and when. In this thesis, we have focused on the decentralized coordination.

We have compared decentralized coordination schemes with respect to three main criteria: 1) efficiency, i.e. how often is each resource accessed by one agent only; 2) fairness, i.e. whether all the agents can access some resource roughly equally often; and 3) rationality, that is whether agents who try to maximize their own utility have an incentive to follow the prescribed coordination scheme.

We have analyzed two different settings for decentralized resource allocation: 1) cooperative, where the agents do not optimize their own utility *per se*, and follow the prescribed protocol; and 2) non-cooperative, where each agent acts to maximize her own utility, without any regard to the utility of the other agents. In the cooperative setting, we have only tried to achieve efficiency and fairness, whereas in the non-cooperative setting, we have focused on rationality as well.

Cooperative Resource Allocation

In this thesis, we have designed a novel approach to achieve a fair and efficient allocation in the cooperative resource allocation problem. We have assumed that the agents access the resources repeatedly, in slots. At the beginning of each slot, all the agents observe a global coordination signal, on which they condition their strategy for that slot. In practice, this coordination signal can be implemented using common clock, radio broadcasting, by observing the decimal part of a specified stock price, etc. The agents learn a different resource allocation for each signal value. Since for each signal, the resource allocation is efficient, the overall resource allocation is efficient too. This is a marked improvement over existing protocols such as ALOHA, that rely on random access with back-off, and that can only achieve a throughput of 37%.

When the resource allocation problem is modeled as a game, the ideal outcome is a correlated equilibrium. The canonical definition of correlated equilibrium assumes a "smart" correlation device, that then tells the "stupid" agents which action to play. In contrast, our solutions assumes "stupid" signal that is not specific to the game, and "smart" agents, who use this signal to learn their strategy.

To learn an efficient allocation, we have proposed a simple algorithm based on randomized back-off. The agents start accessing a randomly chosen resource, and if they collided, in the next round of the game they yield (not access any resource) with a fixed probability. On the other hand, an agent who yielded monitored a randomly chosen resource, and if this was free, the agent accesses it in the next round. This algorithm has an advantage that the agents only need to observe the state of one resource: either the resource they access, or the resource they monitor when they yield. They also need only binary feedback – whether the resource was occupied or not, and whether they collided or not.

We have shown that for an N-agent, C-resource allocation game, our algorithm converges to an efficient allocation in polynomial time with respect to N and C. Since the algorithm is randomized, and since for each coordination signal, the strategies of the agents are independent, the resulting allocation for each signal value is randomly drawn from the space of all efficient allocations. The fairness of this allocation, as measured by *Jain index*, improves as the number of signals K increases.

We have experimentally evaluated our algorithm. We have proposed several modifications of our back-off algorithm, where the back-off probability depends on the number of signals for which the agent already accesses some resource. That way, the agents who access less often back off with lower probability than agents who access more often. This way, they reach an allocation that is more fair.

We have also empirically analyzed the performance of our learning algorithm in case the agent population is dynamic. This is very common for example in wireless networks. We have looked at four scenarios: agents who join later (this corresponds to new nodes joining a wireless

network), agents who restart themselves and start learning from scratch, agents who observe noisy coordination signal, and agents who receive noisy feedback about the occupancy of a resource they accessed.

When new agents join the network, our algorithm is still able to reach an efficient allocation. However, if the new agents are very greedy (and they attempt to access a resource for every signal value), the resulting resource allocation will be unfair. When agents restart their strategies, a strategy that converges faster can have higher efficiency than a more complex, albeit more fair strategy. When the agents receive noisy feedback or noisy coordination signal, our learning algorithm is still able to reach efficient and fair allocation, provided the noisy is low.

Finally, we have compared our learning algorithm to generic multi-agent learning algorithms. These generic algorithms are based on the idea of regret minimization (that is, the minimization of the ex-post difference between the actual payoff and a payoff that the agent could have obtained had she used a different, simple strategy). These regret-based algorithms have been theoretically proven to converge to a distribution of play that is close to a correlated equilibrium. However, when applied to the resource allocation problem, they converged in vast majority of experiments to a single resource allocation. Therefore, their allocation, while efficient, was very unfair compared to the allocation achieved by our algorithm. At the same time, their convergence to an efficient allocation was much slower than for our single-purpose algorithm.

Non-cooperative Resource Allocation

While our coordination scheme for the cooperative resource allocation was able to achieve efficient and fair allocation, it was not rational for self-interested agents to adopt it; such a self-interested agent could access a resource all the time, and make everyone else back off. This is sometimes called the "watch out I am crazy" or "bully" strategy (Littman and Stone (2002)).

To design a rational resource allocation scheme, we have considered the infinitely repeated resource allocation game, where the agents discount future payoffs with a common discount factor $0 < \delta < 1$. Unlike in the cooperative case, the agents can observe the use of all the resources at the same time. Since the game is symmetric and we have assumed that the agents are all identical, we have looked for symmetric subgame-perfect equilibria of the repeated game.

As a measure of how efficient symmetric equilibria are compared to asymmetric ones, we have proposed the price of anonymity. The price of anonymity of a symmetric strategy profile is the ratio between its social payoff (sum of payoffs to each agent when agents adopt the strategy profile) and the maximal social payoff obtained by any asymmetric strategy profile.

We have proposed a symmetric equilibrium based on the ideas of Bhaskar (2000) and Kuzmics

et al. (2010): the agents start by playing randomly, and once they reach a pure-strategy Nash equilibrium of the symmetric game, they adopt a *convention* that prescribes their play from then on. We define the convention formally for any symmetric game as a mapping from the set of pure-strategy Nash equilibria to a vector of continuation strategies for the players. For games where the agents can observe a global coordination signal (like in our cooperative solution above), we have defined an *augmented* convention, that maps a pure strategy NE for every signal value to a vector of continuation strategies for every signal value. We have defined an equilibrium convention as a convention where the continuation strategies are subgame-perfect equilibria of the continuation game.

For the repeated resource allocation game, we have showed that for any (augmented) equilibrium convention, there exists an equilibrium *implementation* – that is, a randomized strategy, that prescribes the players how to play before they reach a pure-strategy NE, so that when taken as a whole the whole strategy is a subgame-perfect equilibrium.

We have then presented two examples of a convention for the resource allocation game – 1) bourgeois, and 2) market convention. In the bourgeois convention, an agent who successfully accessed a resource without collision keeps accessing the same resource forever after. Once the agents play some pure-strategy Nash equilibrium of the game, they will keep playing that NE forever. We have showed that when the number of resources C is low relative to the number of agents N, the expected social payoff in the equilibrium is zero – the price of anonymity is then infinite. This is because some agents can never access any resource successfully, and so they are indifferent between yielding always, and accessing with high probability and risking collisions. For the bourgeois convention, we have shown that the agents can converge to a PSNE in number of steps that is logarithmic in the number of resources C.

To improve the social payoff, we have proposed the market convention. The market convention increases the resource supply by using a coordination signal on which the agents condition their strategy. At the same time, it decreases the demand for resources by charging a price for each successful access. When the agents have common decreasing marginal utility from accessing more often, the price can be set such that they only want to access for one signal value. We have shown experimentally that the market convention has an equilibrium implementation with positive social payoff, and finite price of anonymity.

We have shown that when the agents are cooperative, they can use the coordination signal to achieve a more fair outcome. When the agents are self-interested, using a convention which is more fair leads to higher expected payoff (this was shown by Kuzmics et al. (2010)). But just using the coordination signal is not enough to achieve the fair outcome when the agents are self-interested. Therefore, we needed to introduce the market to limit the demand and to improve social efficiency.

Resources with Non-unit Capacity

In this thesis, we have worked with a model of the resource allocation problem where each resource can only be used by one agent at a time. Here, we will present some of the resource allocation problems where each resource has a fixed capacity of how many agents can use it at the same time. We will give an intuition on how our resources can be apply in such problems. The proper analysis remains future work.

A well-known model of resource allocation where each resource has a fixed capacity is the *El-Farol bar problem* (Arthur (1994)). In this problem, N people simultaneously decide if they should go to a bar or stay home. The bar has a limited capacity C. If there are at most C people in the bar, everyone in the bar is having a good time and prefers to be in the bar rather than be at home. But if there are more than C people in the bar, they would prefer to stay at home instead. The problem can be generalized to a situation where there are multiple bars in which the people can go, each with a different capacity.

Most of the techniques and insights from our work can be generalized to the (multiple) El-Farol bar problem. In a symmetric mixed-strategy Nash equilibrium, the agents choose whether to go to the bar or not randomly. Due to the randomness, sometimes the bar will be filled below its capacity (leading to a loss of utility), and sometimes the bar will be overcrowded. However, a pure-strategy Nash equilibrium in which a fixed set of C agents always goes to the bar is not fair.

When the agents are cooperative, they can use a coordination signal such as the day of the week to learn on which days they should each go to the bar. But when they are self-interested, their decisions on each day will be independent of the other days, and they will go to the bar with probability high enough so that everyone is indifferent between going to the bar or staying home. If staying home has a utility 0, the expected payoff in the equilibrium will be zero too. In order to limit the demand, the bar owners can introduce an entry fee, so that when the agents have a decreasing marginal utility from going to the bar multiple times a week, they will prefer to go once and stay home the rest of the week.

The El-Farol bar problem is useful for modeling several multi-agent resource allocation problems. We will mention two of them: the problem of traffic congestion, and the problem of bidding in keyword auctions.

In the traffic congestion problem, imagine N agents who want to go every morning from City A to City B. They can choose between multiple roads which have each a limited capacity. The overall capacity is not enough for all the agents to go from A to B at the same time. So some agents have to stay home and go later. Until the capacity of the road is reached, the agents can drive at the speed limit. But when the capacity is reached and exceeded, traffic jams occur and the road is congested (Flynn et al. (2009) analyze how these “phantom jams” occur when the road is close to its capacity).When the agents are cooperative, they can use the day of the week to coordinate when to go. But when they are self-interested, they will drive with probability

high enough so that they are indifferent between going early and going late.

Increasing the number of available resources (either capacity of the road or building more roads) can help if the demand stays constant (Duranton and Turner (2009) show that actually, the traffic usually rises when road capacity increases). So we need to reduce the demand, for example by charging a price for using the road (this is increasingly used in cities such as London or Stockholm).

Keyword auctions are used by web search engines to sell advertisement on the search result pages. Lahaie et al. (2007) give an overview of the keyword auction problem. Each search page contains multiple advertisement slots. For each search term (a *keyword*), an advertiser has a utility when a user clicks on her ad. The advertiser submits a bid to the search engine, stating how much she is willing to pay for one click. Using our resource allocation game model, we can model the ad slot as a resource and the keyword as a coordination signal. This is because each advertiser can only have one ad displayed at the same time (she can only access one resource), but she can display ads for multiple keywords (she can access some resource for multiple coordination signals).

When many advertisers bid for the same keyword, the price of the slots increases, and their utility decreases. Cooperative advertisers can learn to only bid for certain keywords, so as to improve their utility. But self-interested advertisers will bid with probability high enough so that the others are indifferent between bidding for a keyword and not bidding. Therefore, we need to decrease the demand of the advertisers for more clicks. Naturally, the advertisers may have decreasing marginal utility from receiving more clicks (this is the case for example in the ad auction version of the Trading Agent Competition (Jordan and Wellman (2010))). We may also charge the advertisers a fixed cost for bidding for a keyword.

Other Future Research Directions

For the cooperative resource allocation problem, we have proposed that the agents can condition their strategy on the value of a global coordination signal that everyone can periodically observe. This coordination signal allows the agents to reach a correlated equilibrium that is a convex combination of Nash equilibria (Hart (2005)). One interesting open question is to identify the class of games where desirable equilibria can be obtained as a convex combination of NE.

For the non-cooperative resource allocation problem, we have identified two conventions that differ in their equilibrium payoffs. We have defined a price of anonymity of a strategy profile as a measure of its efficiency, and price of anonymity of a symmetric game as the ratio between the most efficient symmetric and most efficient asymmetric strategy profile. Bhaskar (2000) and Kuzmics et al. (2010) have defined the egalitarian convention as the most efficient convention for the 2-agent, 1-resource allocation game, and the Nash demand game respectively. It remains an open question to find the most efficient symmetric equilibrium

strategy profile in the general N-agent, C-resource game.

To improve the social payoff in the general resource allocation game, we have defined the market convention. This convention relies on additional assumptions: agents have decreasing marginal utility of accessing more often, and there is mechanism that allows them to pay for successful access. Is there a general method to achieve better payoff than the bourgeois convention, that doesn't require the additional assumptions of the market convention?

The model of resource allocation problems which we used in this thesis assumed that each resource can only be used by one agent at the same time. Furthermore, all the agents had the same utility from using each resource. We have already given some intuition on how we can use our results in problems where one resource may be shared by more than one agent. If the agents have different utilities for using the resources, the game becomes asymmetric. Nevertheless, it might be still important to focus on symmetric equilibria: The main problem in our problem was not that the equilibria were asymmetric, but that there were multiple asymmetric equilibria, and the agents didn't know which one they should play. From a set of multiple *symmetric* equilibria, a convention can easily choose one which the agents will play.

Recently, a new class of strategies for infinitely repeated games called *zero-determinant* strategies has been proposed by Press and Dyson (2012). These strategies are Markovian, in a sense that the agents only base their decision on the outcome of the last round. Instead of considering the expected *discounted* payoff, the authors propose to analyze the payoffs in a *stationary state* of the Markov chain generated by a given strategy vector. A stationary state is a state to which a certain class of Markov chains is guaranteed to converge, irrespective of the initial state. It would be interesting to analyze the symmetric equilibria of the resource allocation game in this new payoff model.

Bibliography

Norman Abramson. The ALOHA system: another alternative for computer communications. In *Proceedings of the November 17-19, 1970, fall joint computer conference*, AFIPS '70 (Fall), pages 281–285, New York, NY, USA, 1970. ACM. doi: 10.1145/1478462.1478502. URL http://dx.doi.org/10.1145/1478462.1478502.

M. S. Alam and A. Z. M. E. Hossain. Throughput analysis of a multichannel slotted-ALOHA protocol in short-haul communication environment for an exponential backoff retransmission scheme. In *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat. No.97TH8237)*, pages 1034–1038. IEEE, 1997. ISBN 0-7803-3676-3. doi: 10.1109/ICICS.1997.652138. URL http://dx.doi.org/10.1109/ICICS.1997.652138.

W.B. Arthur. Inductive reasoning and bounded rationality. *The American economic review*, 84 (2):406–411, 1994.

R. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, March 1974. ISSN 03044068. doi: 10.1016/0304-4068(74)90037-8.

Gabriel Balan, Dana Richards, and Sean Luke. Long-term fairness with bounded worst-case losses. *Autonomous Agents and Multi-Agent Systems*, 22(1):43–63, January 2011. ISSN 1387-2532. doi: 10.1007/s10458-009-9106-9. URL http://dx.doi.org/10.1007/s10458-009-9106-9.

V. Bhaskar. Egalitarianism and efficiency in repeated symmetric games. *Games and Economic Behavior*, 32(2):247–262, August 2000. ISSN 08998256. doi: 10.1006/game.2000.0810. URL http://dx.doi.org/10.1006/game.2000.0810.

A. Blum and Y. Mansour. Algorithmic game theory. In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, editors, *Algorithmic Game Theory*, chapter 4. Cambridge University Press, September 2007.

Andrea Cavagna. Irrelevance of memory in the minority game. *Physical Review E*, 59(4): R3783–R3786, April 1999. doi: 10.1103/PhysRevE.59.R3783.

Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority Games: Interacting Agents in Financial Markets (Oxford Finance)*. Oxford University Press, New York, NY, USA, January 2005. ISBN 0198566409.

Xi Chen and Xiaotie Deng. Settling the complexity of Two-Player nash equilibrium. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 261–272. IEEE, 2006. ISBN 0-7695-2720-5. doi: 10.1109/FOCS.2006.69. URL http://dx.doi.org/10.1109/FOCS.2006.69.

S. Cheng, A. Raja, L. Xie, and I. Howitt. A distributed constraint optimization algorithm for dynamic load balancing in wlans. In *The IJCAI-09 Workshop on Distributed Constraint Reasoning (DCR)*, 2009.

K. Decker and J. Li. Coordinated hospital patient scheduling. In *Multi Agent Systems, 1998. Proceedings. International Conference on*, pages 104–111. IEEE, 1998.

Gilles Duranton and Matthew A. Turner. The fundamental law of road congestion: Evidence from US cities. *National Bureau of Economic Research Working Paper Series*, pages 15376+, September 2009. URL http://www.nber.org/papers/w15376.

William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition*. Wiley, 3 edition, January 1968. ISBN 0471257087.

M. R. Flynn, A. R. Kasimov, J. C. Nave, R. R. Rosales, and B. Seibold. Self-sustained nonlinear waves in traffic flow. *Physical Review E*, 79(5):056113+, May 2009. doi: 10.1103/PhysRevE.79.056113. URL http://dx.doi.org/10.1103/PhysRevE.79.056113.

Dean P. Foster and Rakesh V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40–55, October 1997. ISSN 08998256. doi: 10.1006/game.1997.0595.

Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, May 1986. ISSN 00129682. doi: 10.2307/1911307. URL http://dx.doi.org/10.2307/1911307.

Drew Fudenberg and Jean Tirole. *Game Theory*. The MIT Press, August 1991. ISBN 0262061414. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262061414.

N. Gast. Computing hitting times via fluid approximation: application to the coupon collector problem. *Arxiv preprint arXiv:1107.3385*, July 2011.

Trond Grenager, Rob Powers, and Yoav Shoham. Dispersion games: general definitions and some specific learning results. In *Proceedings of the Eighteenth national conference on Artificial intelligence (AAAI-02)*, pages 398–403, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence. ISBN 0-262-51129-0.

Sergiu Hart. Adaptive heuristics. *Econometrica*, 73(5):1401–1430, 2005. ISSN 00129682. doi: 10.2307/3598879. URL http://dx.doi.org/10.2307/3598879.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, September 2000. ISSN 00129682. doi: 10.2307/2999445.

Rajendra K. Jain, Dah-Ming W. Chiu, and William R. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical report, Digital Equipment Corporation, September 1984.

J.L.W.V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

P.R. Jordan and M.P. Wellman. Designing an ad auctions game for the trading agent competition. *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 147–162, 2010.

Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th annual conference on Theoretical aspects of computer science*, STACS'99, pages 404–413, Berlin, Heidelberg, 1999. Springer-Verlag. ISBN 3-540-65691-X. URL http://portal.acm.org/citation.cfm?id=1764944.

C. Kuzmics, T. Palfrey, and B. Rogers. Symmetric players in repeated games: Theory and evidence, 2010.

S. Lahaie, D.M. Pennock, A. Saberi, and R.V. Vohra. Sponsored search auctions. *Algorithmic Game Theory*, pages 699–716, 2007.

Kevin Leyton-Brown and Yoav Shoham. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*. Morgan & Claypool, San Rafael, CA, 2008.

Michael Littman and Peter Stone. Implicit negotiation in repeated games intelligent agents VIII. In John-Jules Meyer and Milind Tambe, editors, *Intelligent Agents VIII*, volume 2333 of *Lecture Notes in Computer Science*, chapter 29, pages 393–404. Springer Berlin / Heidelberg, Berlin, Heidelberg, June 2002. ISBN 978-3-540-43858-8. doi: 10.1007/3-540-45448-9_29. URL http://dx.doi.org/10.1007/3-540-45448-9_29.

P. Mahonen and M. Petrova. Minority game for cognitive radios: Cooperating without cooperation. *Physical Communication*, 1(2):94–102, June 2008. ISSN 18744907. doi: 10.1016/j.phycom.2008.03.001.

Pragnesh J. Modi, Hyuckchul Jung, Milind Tambe, Wei-Min Shen, and Shriniwas Kulkarni. Dynamic distributed resource allocation: A distributed constraint satisfaction approach intelligent agents VIII. In John-Jules C. Meyer and Milind Tambe, editors, *Intelligent Agents VIII*, volume 2333 of *Lecture Notes in Computer Science*, chapter 19, pages 264–276. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2002. ISBN 978-3-540-43858-8. doi: 10.1007/3-540-45448-9_19. URL http://dx.doi.org/10.1007/3-540-45448-9_19.

D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, pages 124–143, May 1996. ISSN 0899-8256.

Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, September 1997. ISBN 0674341163.

John Nash. Non-Cooperative games. *The Annals of Mathematics*, 54(2):286–295, September 1951. ISSN 0003486X. doi: 10.2307/1969529. URL http://dx.doi.org/10.2307/1969529.

J. R. Norris. *Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, July 1998. ISBN 0521633966.

Christos H. Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *Journal of the ACM*, 55(3):1–29, July 2008. ISSN 0004-5411. doi: 10.1145/1379759.1379762. URL http://dx.doi.org/10.1145/1379759.1379762.

William H. Press and Freeman J. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, June 2012. ISSN 1091-6490. doi: 10.1073/pnas.1206569109. URL http://dx.doi.org/10.1073/pnas.1206569109.

Martin Raab and Angelika Steger. *"Balls into Bins" — A Simple and Tight Analysis*, volume 1518 of *Lecture Notes in Computer Science*, chapter 13, pages 159–170. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. ISBN 978-3-540-65142-0. doi: 10.1007/3-540-49543-6_13. URL http://dx.doi.org/10.1007/3-540-49543-6_13.

Vernon Rego. Naive asymptotics for hitting time bounds in markov chains. *Acta Informatica*, 29(6):579–594, June 1992. ISSN 0001-5903. doi: 10.1007/BF01185562.

R.M. Richman. Recursive binary sequences of differences. *Complex Systems*, 13(4):381–392, 2001.

Robert Savit, Radu Manuca, and Rick Riolo. Adaptive competition, market efficiency, and phase transitions. *Physical Review Letters*, 82(10):2203–2206, March 1999. doi: 10.1103/PhysRevLett.82.2203.

Jeffrey Shneidman, Chaki Ng, David C. Parkes, Alvin Auyoung, Alex C. Snoeren, Amin Vahdat, and Brent Chun. Why markets could (but don't currently) solve resource allocation problems in systems. In *In USENIX '05: Proceedings of the 10th USENIX Workshop on Hot Topics in Operating Systems*, page 7, 2005.

Olga Taussky. A recurring theorem on determinants. *The American Mathematical Monthly*, 56 (10):672–676, 1949. ISSN 00029890. URL http://www.jstor.org/stable/2305561.

Katja Verbeeck, Ann Nowé, Johan Parent, and Karl Tuyls. Exploring selfish reinforcement learning in repeated games with stochastic rewards. *Autonomous Agents and Multi-Agent Systems*, 14(3):239–269, June 2007. ISSN 1387-2532. doi: 10.1007/s10458-006-9007-0. URL http://dx.doi.org/10.1007/s10458-006-9007-0.

Luděk Cigler

lcigler@gmail.com
Avenue de Morges 76, 1004 Lausanne, Switzerland
Nationality: Czech Republic, EU
Tel.: +41 76 596 11 24
Date of birth: 24.11.1983

Education

PhD Candidate, Computer Science – Artificial Intelligence 2008–2012
(expected)
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Grade average: 6.00 (best grade 6.0)

Research internship 06/2011–07/2011
Harvard University, Cambridge, CA, USA

MSc, Theoretical Computer Science 2006–2008
Charles University, Prague, Czech Republic
Grade average: 1.00 (best grade 1), first class

Erasmus exchange, Artificial Intelligence 09/2006–02/2007
University of Amsterdam, Amsterdam, The Netherlands
Grade average: 8.7 (best grade 10.0)

BS, Computer science 2003–2006
Charles University, Prague, Czech Republic
Grade average: 1.00 (best grade 1), first class

Work Experience

RadicalSoft Ltd., Jersey, UK (remotely, 10 hours per month) 07/2009–12/2010
Consultant, SW development:
Designed and implemented AwesomeSearch, an extended search bar for Firefox. Developed parts of FireTorrent, a BitTorrent client for Firefox.

AllPeers Ltd., Prague, Czech Republic, (40% contract) 03/2007–03/2008
Internship, SW development:
Developed new features for a Firefox P2P file-sharing extension, one of the largest Firefox extensions at that time.

Powel ASA, Trondheim, Norway 07/2004–09/2004
Internship, SW development:
Designed and developed automated regression tests for a web application used for monitoring energy consumption. Improved the UI of that web application to follow HTML/CSS standards.

IT Skills

Main: Python, C++, Matlab, JavaScript, HTML/CSS, Mac OS X, Linux
Familiar: Java, SQL (MySQL, PostgreSQL), PHP, Haskell, Prolog

Language Skills

Czech (native), **English** (native-like fluency, C2 level), **French** (professional fluency, C1 level), **German** (basic fluency, B2 level), **Spanish** (basic fluency, B2 level)

Activities

Participated in several international volunteer workcamps on environmental issues (Denmark, Sweden, Italy)
Track&field coach for a group of 13 to 15 year olds (2004–2008)
Venture Challenge, course on entrepreneurship, HEC Lausanne, Spring 2011

List of Personal Publications

L. Cigler and B. Faltings. Reaching correlated equilibria through multi-agent learning. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 509–516. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

L. Cigler and B. Faltings. Symmetric subgame perfect equilibria for resource allocation. In *Proceedings of the 26th national conference on Artificial intelligence (AAAI-12)*, Menlo Park, CA, USA. American Association for Artificial Intelligence, 2012.