Institut für Lebensmittel- und Ressourcenökonomik

Spatial competition of learning agents in agricultural procurement markets

D i s s e r t a t i o n

zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr.agr.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

Bonn

vorgelegt von

Hamed Khalili

aus

Karaj, Iran

Bonn 2019

Referent:	Prof. Dr. Thomas Heckelei
Korreferent:	Prof. Dr. Alfons Balmann

Tag der mündlichen Prüfung: 04.03.2019

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität

Acknowledgments

My supervisor Professor Thomas Heckelei gave me the great opportunity at ILR, I gained a lot of insights from him at various points of my research and his awesome academic knowledge was inspiring me through the whole time in the Ph.D. pursuit. Thank you very much.

I would like to thank Professor Kathy Baylis from University of Illinois Urbana-Champaign for her comments and suggestions on the final version of the thesis. I gratefully appreciate brilliant support provided by Christine Wieck and Wolfgang Britz, who encouraged and guided me through initiating phases of this research. Feedbacks proposed by Marten Graubner and Sebastian Rasch in the course of my project have been illuminating. Thank you for helpful remarks! Special mention goes to Professor Mathias Erlei from Clausthal University of technology for nurturing my enthusiasm for Economics. I would also like to acknowledge the funding source of my Ph.D., which was German Science Foundation. I appreciate the evaluation of my thesis by Professor Alfons Balmann from Leibniz Institute of Agricultural Development in Transition Economies.

I am indebted to my wonderful mother and my father. I love you two forever! My lovely wife Zahra has given selflessly so much love and care. She was bearing a lot of ups and downs of my Ph.D. time. I love you very much. You are amazing!

Kurzfassung

Räumlich verteilte Betriebe liefern Rohmilch als primären Input an eine kleine Anzahl milchverarbeitender Großbetriebe. Der räumliche Wettbewerb der verarbeitenden Betriebe hat kurz- bis langfristige Auswirkungen auf die Farm- und Molkereistruktur, da er die regionale Nachfrage nach Rohmilch sowie den daraus resultierenden Rohmilchpreis bestimmt. Eine Reihe neuerer analytischer und empirischer Beiträge analysieren den räumlichen Preiswettbewerb von Verarbeitungsunternehmen auf Milchmärkten. Agentenbasierte Modelle (ABM) werden neuerdings als ‚Bottom-up-Ansätze' eingesetzt, um die emergenten Marktergebnisse von autonom entscheidenden und interagierenden Marktakteuren besser zu verstehen. Trotz der Stärken von ABMs ist die Berücksichtigung interaktiven Lernens durch intelligente Agenten in ABMs nicht ausreichend gereift. Obwohl die Literatur von Multi-Agenten-Systemen (MASs) und Multi-Agenten-basierte Simulation ökonomischer Interaktionsprozesse eigentlich verwandte Forschungsgebiete sind, haben sie bisher weitgehend getrennte Wege beschritten. Diese Dissertation trägt zur Entwicklung der Grundlagen für das Design von lernenden Agenten in räumlichen ökonomischen ABMs bei. Jedes der drei Hauptkapitel der Arbeit untersucht ein zentrales Thema für die Gestaltung interaktiver Lernsysteme mit dem übergeordneten Ziel, das Entstehen von Preisverhalten in realen räumlichen Agrarmärkten besser zu verstehen.

Ein wichtiges Problem in der Literatur zum räumlichen Wettbewerb ist die lückenhafte theoretische Erklärung für das beobachtete Kartellverhalten in oligopsonistischen Märkten. Das erste Hauptkapitel leitet theoretisch ab, wie die Einbeziehung von Voraussicht in die Preispolitik von Agenten in räumlichen Märkten das System in Richtung kooperativer Nash-Gleichgewichte bringen kann. Es wird gezeigt, dass die Berücksichtigung eines einfachen Maßes an Voraussicht den Agenten die Möglichkeit eröffnet ansonsten endlose Preiskriege zu beenden.

Die Einführung einer „outside Option“ in die Entscheidungsmöglichkeit der Agenten im Rahmen eines dynamischen Preisspiels zeigt die Korrelation negativer Grenzerträge strategischen Kalküls mit der Relevanz der Transportkosten.

Im zweiten Hauptkapitel stellen wir einen neuen Lernalgorithmus für rationale Agenten vor, der die Methode eines „hierarchischen Gradienten“ (H-PHC) verwendet. Während MASs-Algorithmen typischerweise nur auf kleine Probleme anwendbar sind, zeigen wir experimentell, wie multiple rationale H-PHC Agenten in der Lage sind, das Koordinationsproblem in einer Vielzahl von räumlichen (und nicht-räumlichen) Marktspielen, charakterisiert durch große Entscheidungsräume, mit mäßigem Rechenaufwand zu überwinden.

Die theoretische Erklärung von Preisgleichgewichten in räumlichen Märkten ist in der Literatur umstritten. Die Mehrheit der Artikel aber erklärt das Preisverhalten (Molkereipreis und Frachtabsorption) allein mit der räumlichen Struktur der Märkte. Basierend auf einem computergestützten Ansatz mit interaktiv lernenden Agenten im zweidimensionalen Raum, schlägt das dritte Hauptkapitel vor, dass die Erklärung des Umfangs der Frachtabsorption allein mit dem Faktor Raum unvollständig ist. Das Preisverhalten landwirtschaftlicher Verarbeiter, speziell ihre Fähigkeit zur Koordination und Erzielung von gegenseitig vorteilhaften Ergebnissen, ist zusätzlich abhängig von ihrer Fähigkeit voneinander zu lernen.

Schlüsselwörter: *Räumliche Agrarmärkte, agentenbasierte Modellierung, oligopsony, Preisbildung*

Abstract

Spatially dispersed farmers supply raw milk as the primary input to a small number of large dairy-processing firms. The spatial competition of processing firms has short- to long-term repercussions on farm and processor structure, as it determines the regional demand for raw milk and the resulting raw milk price. A number of recent analytical and empirical contributions in the literature analyse the spatial price competition of processing firms in milk markets. Agent-based models (ABMs) serve by now as computational laboratories in many social science and interdisciplinary fields and are recently also introduced as bottom-up approaches to help understand market outcomes emerging from autonomously deciding and interacting agents. Despite ABMs' strengths, the inclusion of interactive learning by intelligent agents is not sufficiently matured. Although the literature of multi-agent systems (MASs) and multi-agent economic simulation are related fields of research they have progressed along separate paths. This thesis takes us through some basic steps involved in developing a theoretical basis for designing multi-agent learning in spatial economic ABMs. Each of the three main chapters of the thesis investigates a core issue for designing interactive learning systems with the overarching aim of better understanding the emergence of pricing behaviour in real, spatial agricultural markets.

An important problem in the competitive spatial economics literature is the lack of a rigorous theoretical explanation for observed collusive behavior in oligopsonistic markets. The first main chapter theoretically derives how the incorporation of foresight in agents' pricing policy in spatial markets might move the system towards cooperative Nash equilibria. It is shown that a basic level of foresight invites competing firms to cease limitless price wars. Introducing the concept of an outside option into the agents' decisions within a dynamic pricing game reveals

how decreasing returns for increasing strategic thinking correlates with the relevance of transportation costs.

In the second main chapter, we introduce a new learning algorithm for rational agents using H-PHC (hierarchical policy hill climbing) in spatial markets. While MASs algorithms are typically just applicable to small problems, we show experimentally how a community of multiple rational agents is able to overcome the coordination problem in a variety of spatial (and non-spatial) market games of rich decision spaces with modest computational effort.

The theoretical explanation of emerging price equilibria in spatial markets is much disputed in the literature. The majority of papers attribute the pricing behavior of processing firms (mill price and freight absorption) merely to the spatial structure of markets. Based on a computational approach with interactive learning agents in two-dimensional space, the third main chapter suggests that associating the extent of freight absorption just with the factor space can be ambiguous. In addition, the pricing behavior of agricultural processors – namely the ability to coordinate and achieve mutually beneficial outcomes - also depends on their ability to learn from each other.

Keywords: *spatial agricultural markets, agent-based modelling, oligopsony, pricing*

Contents

List of Tables

List of Figures

Abbreviations

ABM	Agent based model
AI	Artificial intelligence
ANN	Artificial neural network
COOP	Cooperative Firm
CP	Competitive pricing
EU	European union
FOB	Free on board pricing
HS	Hotelling-Smithies conjecture
IOF	Investor owned firm
MARL	Multi-agent reinforcement learning
MAS	Multi-agent system
MDP	Markov decision processes
NCP	Non-competitive pricing
NE	Nash equilibrium
OD	Optimal discriminatory pricing
PD	Prisoner's dilemma game
PM	Price matching conjecture
RL	Reinforcement learning
UD	Uniform delivered pricing

Chapter 1
Introduction

1.1 Overall research background and objectives

Although microeconomic textbooks introduce agricultural product markets as examples for perfectly competitive markets, a large number of studies emphasize that in reality such markets show oligopsonistic structures, especially in light of dramatically increased concentration in food processing (Sexton, 1990 and 2012). A multitude of spatially dispersed suppliers, relatively few processors and costly to transport raw products often also characterize the raw milk market. Collusive behavior on the part of processors has been studied in raw milk procurement markets quite recently (Graubner et al., 2011a; Huber, 2009; Bundeskartellamt 2009; Huber, 2007 a; Huck et al., 2006; Alvarez et al., 2000). The bulky raw milk product can be delivered only to limited number of buyer locations. In addition, the perishability of raw milk weakens the dairy farmers' bargaining position in negotiations with dairy processors.

The nature of competition at the processor stage has short- to long-term repercussions on the structure of dairy farming, as it determines the regional demand for raw milk but also the raw milk prices dairy processors are willing to pay.

Most spatial competition models in the literature rely on inadequate analytical assumptions. For example, the theoretical studies of Tribl (2012) and Koller (2012) assume simplified spatial shapes of market, typically linear, one-dimensional markets with farmers distributed uniformly along a straight line and

processors located at either endpoints of the market. Although these models present strong analytical basics, they fail to consider complex environment features like dynamic interactions, two-dimensional market shapes, asymmetric pricing policies of firms or pricing policies with various degrees of freight absorptions.

An additional drawback of theoretical spatial competition models is to assume that objectives and corresponding rational reasoning of agents is common knowledge (Binmore, 1987). In a real market, however, agents make decisions without exact knowledge about the multiple key drivers of their market environment, e.g. other agents' payoffs and preferences.

Computational economic models, especially ABMs are recently proposed to cope with some of the deficiencies mentioned above. These frameworks are able to simulate actions and interaction of *autonomous* agents in complex environments (Grimm and Railsback, 2005). The emergence of positive and negative effects following policy and non-policy shocks may be detected by simulating decisions of many interdependent agents. However, individual agents must be equipped with appropriate adaptive decision mechanisms to successfully simulate such emergent behaviour at the system level (Kirman, 2011).

In spatial competition context, each processor agent needs to dynamically keep up with the changes in the behavior of other agents. Finding the optimal pricing policy of processor agents among multiple, strategically interacting agents is a complex task. Decision rules are supposed to be continuously under review and revision. This problem is addressed as one of a "moving target" problem in the literature of MASs (Vidal, 1998). It is conceivable that agents obeying unpretentious rules -who behave in a non-learning fashion-, typically fail to accomplish desired outcomes in strategic systems. To fair better, agents must acquire knowledge about environment and other agents through the course of interaction, i.e. by *Learning.*

The realm of Multi-Agent-Systems (MASs) is the most noticeable research area addressing efficient approaches in strategic learning problems. Most studies in the MASs literature suggest *Reinforcement learning* as the relevant method used by learning agents inhibiting interactive worlds.[1] On the other hand although MASs and multi agent economic simulation are related fields of research they have progressed along separate paths. While MASs learning algorithms are predominantly applicable to just small problems (Busoniu et al., 2010), computational economics requires learning methods, which are able to overcome the adaptation problem in market games with rich decision spaces.

Many studies in the computational economics literature (e.g. the study of Graubner et al., 2011 in spatial competition) employ evolutionary algorithms to understand the equilibrium behaviour. A number of arguments support the use of genetic algorithms: "… *if we are not concerned with the exact details of individual learning processes, evolutionary algorithms are a sufficient tool to model learning processes on a population level* (Brenner, 2005, p.39)." Evolutionary algorithms can be quite useful for some classes of complex problems especially when we are not concerned with the detailed dynamics and learning properties of the system. However, interpreting the dynamics of genetic algorithms as individual learning processes seems not appropriate (Brenner, 2005, P.39). Mostly evolutionary approach might not compel Agent-based system designers to investigate precise clues with respect to individual learning *mechanism* of system inhabitants. Given these issues Brenner (2005) believes that *"… it is surprising that the use of genetic algorithms, and especially the original genetic algorithms, has widely spread in simulating economic learning processes* (Brenner, 2005, P.39)."

[1] The new introduced learning algorithm in our thesis, Hierarchical Policy Hill Climbing (H-PHC) can be also categorized in respect to Reinforcement learning methods.

In this thesis we would be more concerned with real humans characteristics -like *foresight in decision, dynamic programming, reinforcement learning* and *hierarchical elaboration of choices*- which might be crucial to understand the economic *behaviour* of *learning* agents in spatial markets.

Each of three main chapters of this thesis investigates one central issue, which might serve as small steps towards designing improved MASs in the context of spatial agricultural markets:

The analytical chapter 2 addresses the issue of Non-existence of a pure strategy Nash equilibrium triggering the possibility of an escalating arms race with no end (Schuler and Hobbs, 1982). We introduce a learning theory analogous to the sequential move adjustment process in Maskin and Tirole (2001) and show that through introduction of foresight into the agents' pricing policy, the cyclic price wars move towards cooperation. Our rigorous results by means of a simplified market setting may hint at the reason why price-matching is the predominant observed behavior of processing firms in many agricultural markets, especially also in raw milk procurement markets. We discuss to what extent implications of our model coincide with prior studies in the literature. Especially we introduce the concept of Outside Option.

In chapter 3 we investigate the issue of two-dimensional and non-asymmetric spatial price formation between farmers and processors in spatial environments by introducing a new Learning algorithm. While Multi-agent reinforcement learning algorithms are predominantly applicable to just small problems (Busoniu et al., 2010), we introduce the H-PHC (hierarchical policy hill climbing) algorithm in our work, which is able to overcome the coordination problem in a variety of spatial (and non-spatial) market games with rich decision spaces. Indeed our investigation in chapter 3 concerns chiefly about the issue of *Scalability*. We show experimentally how a community of multiple rational H-PHC agents perform a *rational* and *convergent* learning process where agents do not need to model each other explicitly as agents. This will lead to extra computational efficiency.

In chapter 4, we implement some learning-based scenarios of pricing in agricultural processor markets in two-dimensional space and in the presence of policies with variable levels of freight absorption.

Most previous studies in the agricultural economics literature[2] attribute the degree of freight absorption by processing firms just to the spatial structure of markets. We compare two opposite poles of learning aptitude of processors. These are low-coordination and high-coordination scenarios. Our interaction scenarios propose that associating the extent of freight absorptions by pricing policy of firms just with factor space in spatial markets might depend on the extent of coordination between firms and hence policy recommendations based on such measures can lead us in the wrong direction. In addition to the spatial structure of the market - the pricing behavior of agricultural processors also depends on their ability to learn from each other.

In the section 1.2 of this chapter we introduce some required background knowledge from literature regarding core theory and tools used in following chapters. In section 1.3 we discuss the summary of contributions, key results and limitations of following chapters.

1.2 **Background Knowledge**

1.2.1 *Economic structure of the dairy market*

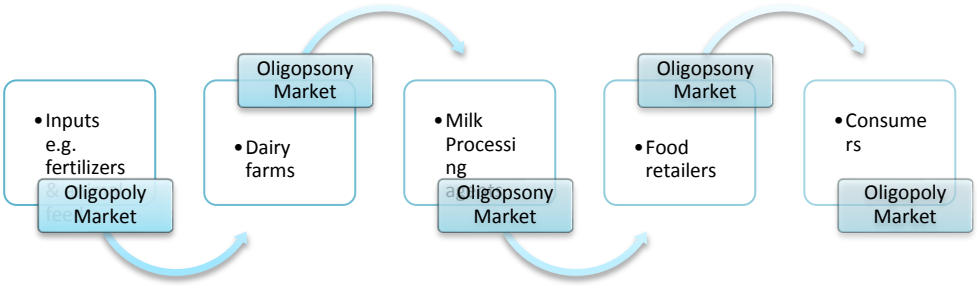
The milk sector is not only receptive to structural changes at farm level but also sensitive to agents' interaction at the dairy processing level. Competition between processors will have in short- to long-term repercussions on the dairy farming structure, as they determine the regional demand for raw milk but also raw milk prices dairy processors are willing to pay. The number of dairy processors and their operating sites in Germany show a declining trend for years and this has been observed in most other EU countries as well (Boysen and Schröder, 2006).[3] Although agricultural markets often serve as examples for competitive markets in

[2] See e.g. Zhang and Sexton (2001).

[3] Since 1984, the dairy production sector in Europe was regulated by a milk quota system aiming at controlling milk production and price while limiting public expenditure. The liberalization of the EU milk market respectively the abolishment of the quota system in 2015 might have caused this trend to speed up.

microeconomics textbooks, in reality many agricultural procurement markets are characterized by an oligopsonistic economic structure (Sexton, 2012). Tribl and Salhofer (2013) offer an empirical and theoretical survey regarding the market power along the dairy supply chain.

Figure 1.1: Stylized demonstration of dairy supply chain based on Tribl and Salhofer (2013).



In order to capture the potential power relationships in each stage of the supply chain depicted in Figure 1.1, one must consider the options of each player to leave ongoing negotiations if its outside gain is higher than that of current bargaining process. These *Outside Options* (Osborne and Rubinstein, 1990) may grant different market players to have different market powers. For example, retailers might be able to swap dairy product suppliers. Likewise, the most important Outside Option of a dairy farm is switching to another dairy processor firm. Obviously, the availability and attractiveness of the Outside Options determines the credibility of the threat to abort negotiations and thus to exert power in the negotiations. It follows that if one side has more attractive Outside Options it possesses greater bargaining power and hence, will be less willing to accept a lower pay-off.

Intuitively, a precise analysis of price and income effects in the dairy industry requires a comprehensive overview of price transmissions along the supply

chain.[4] The emphasis in this thesis is however limited to examining the outcome of the interaction between farm and dairy processing level. The literature offers considerable empirical evidence on collusive behavior of processors towards farms in the raw milk market (Graubner et al., 2011a; Huber, 2009; Bundeskartellamt, 2009; Huber, 2007 a; Huck et al., 2006; Alvarez et al., 2000). The above mentioned concentration and move to oligopsonistic structures on the dairy processors side is accompanied by changes in land use and more generally long-term development of farms and might cause wide distributional effects at farm level. Concerns regarding market and or bargaining power in the dairy supply chain market has led to a recent sector inquiry by the German Federal Cartel Agency (2012). Hence, an improved understanding of farm structural changes requires the analysis of dairy firms' competition in raw milk markets.

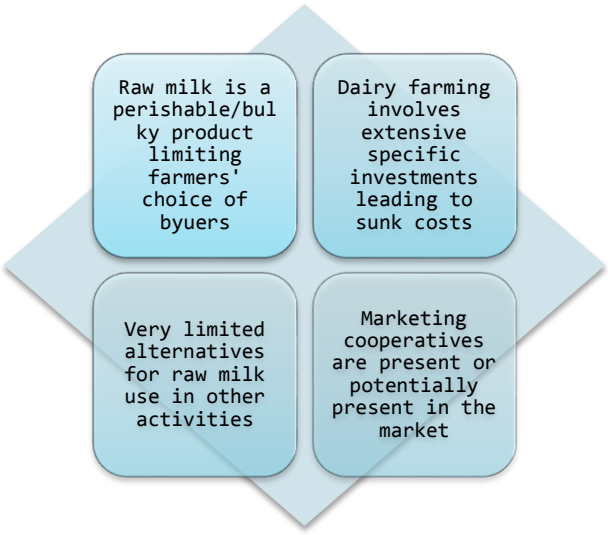
1.2.2 *Contractual system in Dairy markets*

The production of agricultural raw products has some unique characteristics that distinguish it from other commodity markets. Raw milk production can be seen as an important example of agricultural markets showing four distinctive characteristics (depicted in Figure 1.2 following Rogers and Sexton, 1994). Many investments are relationship-specific. Milk suppliers customize their equipment and invest in production facilities that suit particular needs of their dairy business partners. Once built, buildings and machinery are difficult to use differently and can be considered as sunk cost. For example, a milking parlor would be of no use for other production activities and the barn needs reconditioning to suit other production activities. Commercial vehicles on the other had can more easily be sold and used by another firm in a different industry (Wieck and Mosnier, 2011).[5]

[4] Another empirical study regarding pricing along the supply chain is done by Hellberg-Bahr et al. (2010).

[5] Basically capital investments are sunk (specific) when the unit value of investment is greater than the unit value of disinvestment.

Figure 1.2: Distinctive structural dimensions of the raw milk market in line with Rogers and Sexton (1994).



The perishability of raw milk and associated time constraint renders a weak bargaining position for farmers at the negotiation stage with dairy processors. Having the lower discount rate as consequences of perishability may cause a *Hold-up* situation in relationship of dairy processors with producing farms (Osborne and Rubinstein, 1990; Binmore et al., 1986). Sunk costs associated with a farm's specific investment is not compatible with the opportunistic behavior of dairy processing firms without contractual or property rights arrangements. Laffont and Tirole (1993) mention that if contracts are complete in the sense of using all relevant and contractible information, they may provide incentives that minimize efficiency losses. This might rationalize the commonness of conclusion of *contract*s in dairy markets. The German Federal Cartel Agency (BKA, 2009, p.73) reports that a high share of milk is committed in the long-term via supply contracts, i.e. it is not freely available on the spot market.

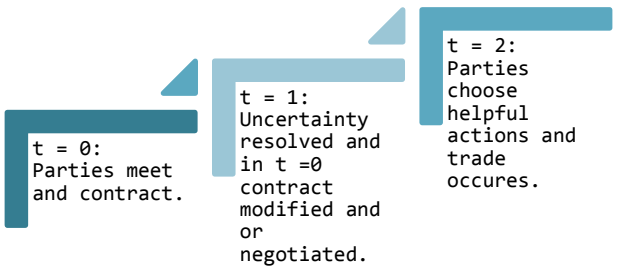
Note that the main feature of market contracts considered in this thesis is the product price. However, decision-makers in real markets include factors other than price into their contracts. Indeed the most important reason for farms to stay with the same processor or switch to another is assumed in our study to be the *delivery price*. Empirical studies of the German milk market reveal that farmers' preferences regarding contract design are more divers. In addition to price,

quantity and quality regulation, contract period, contract termination conditions and some further regulatory issues are relevant attributes of delivery contracts (Steffen et al., 2009).

1.2.3 *Existence of cooperative firms*

Hart (2009) argues that even when originally parties have agreed to cooperate under the terms of a rigid contract that fixes price, parties may withhold cooperation until renegotiation after the resolution of uncertainty lead to changes in costs/values of the contract. This cooperation may conclude a number of "helpful" or "cooperative" actions at some time t=2 (Figure 1.3), which by no means could be anticipated in time 0.

Figure 1.3: Contract timeline based on Hart (2009).



In this sense numerous - perhaps even most - contracts observed in the real world are highly incomplete. A vast literature on vertical integration (Grossman and Hart, 1986) describes how inefficiency might emerge in situations that feature relation-specific investment. This scenario generally results in a trade-off. Specific investments yield a larger surplus to be available for the partners but weaken the ex-post bargaining position of the investor. An inefficient level of under-investment results from the threat of being exploited by the other party in the negotiation stage after the investment took place. Hart and Moore (2008) explain that the risk of being held up might not necessarily be healed through contract. There may be other more prosaic reasons. For example imagine -due to a bad unexpected state of the downstream market- the processor party partner withholds cooperation. Courts may enforce compliance within the *letter* of contract but not the compliance within the *spirit* of contract. The dairy firm can

e.g. make the life difficult for milk producers by quibbling about quality details or by delaying payments due to some strict terms of the contract.

Property right models propose that the under-investment problem can be rectified by means of vertical integration. This might be one possible explanation why Germany dairy farmers are integrated with dairy processing and marketing through dairy cooperatives (COOPs). The theory on COOPs suggests that the right of control over assets should be assigned to agents whose quasi rents are under risk from hold-up behavior of profit maximizing processors (Grossman and Hart 1990). Indeed the costs of contracting generally might increase more than the costs of vertical integration as assets become more specific (Klein et al., 1978). The share of German COOPs in milk processing is rather high (Spiller 2008b). COOPs process around two-thirds of milk produced in Germany and among the five largest dairies - measured by the quantity of milk processed - there are four COOPs (Steffen et al., p.6). Note that processor agents in our study are assumed to be fully rational, profit-maximizing processors even though several alternative objective functions are suggested specifically for COOPs (Cotterill, 1987).[6]

1.2.4 *Basic interaction model in Oligopsony*

The *conjectural variation approach* is the corner stone of most analytical and empirical approaches studying oligopsonistic markets. Assuming that there are N non-cooperative dairy plants (milk processors), the production function for a homogeneous good of the i-th processor in the area is

$$q_i = f\left(x_i, s_i\right) \tag{1.1}$$

where q_i is the output quantity of the representative commodity of milk products produced by the *i*-th dairy plant, x_i is the corresponding quantity of raw milk

[6] Despite of this there are empirical observations indicating that the behavior of cooperative firms is similar to profit maximizer investor-owned investor owned IOF firms. The question if and to what extend COOPs might maximize profits is a very deep and controversial question that depends heavily on the governance structure and the internal incentive system of the specific organization.

bought and s_i is a vector of other inputs. We assume throughout the thesis that the dairy agents are price takers in the output (down-stream) market as well as in the market for other production inputs, but that they exert market power in the raw milk input market. The raw milk (inverse) supply is expressed as

$$w_X = f^{'}(X, S) \tag{1.2}$$

$X = \sum_{i=1}^{N} x_i$ describes the sum over all milk delivered by the farms, *S* presents a vector of supply shifters and w_X denotes the market price of raw milk. The (short- to medium-run) profit of the *i*-th dairy plant given a certain production capacity is given by

$$\Pi_i = pf(x_i, s_i) - w_X x_i - w_S s_i \tag{1.3}$$

where p is the price for the final processed dairy output (assumed to be a representative commodity) and w_S is a vector of prices for the other production factors. Deriving first order condition maximizing profit Π_i with respect to x_i yields

$$\frac{\partial \Pi_i}{\partial x_i} = p\frac{\partial f(x_i, s_i)}{\partial x_i} - w_X(1 + \varphi_i / \varepsilon) = 0 \tag{1.4}$$

where $\varphi_i = (\partial X/\partial x_i)(x_i/X)$ is the i-th dairy plant's conjectural elasticity in the input market for raw milk and $\varepsilon = (\partial X/\partial w_X)(w_X/X)$ is the market price elasticity of raw milk supply. Rearranging terms leads to the following expression for the raw milk price:

$$w_X = pf_{x_i} / (1 + \varphi_i / \varepsilon) \tag{1.5}$$

$f_{x_i} = \partial f(x_i, s_i)/\partial x_i$ is the marginal product of raw milk input used by the i-th dairy plant. According to the literature (Appelbaum, 1982; Azzam and Pagoulatos, 1990) the factor φ_i offers possibilities econometrically testing market structure. If $\varphi_i = 0$, then the raw milk market is perfectly competitive, i. e. the marginal product of raw milk of each dairy plant equals the market price. If $\varphi_i = 1$, then the market for raw milk is monopsonistic, i.e. the dairy plants act

like a monopsony and consequently the marginal factor cost should be equal to the value marginal product. Intermediate values of φ_i indicate the presence of oligopsonistic market behavior to varying degrees (Perekhozhuk et al., 2013).

1.2.5 *ABMs and MASs*

The complexity of determining price policies of dairy agents in spatial markets with strategic interaction makes them difficult to analyse analytically. The investigation of such complex systems through ABMs is well established as an alternative approach in the literature (Happe et al. 2006, Lobianco et al. 2010; Schreinermachers and Berger, 2011; Ostermeyer et al. 2011). ABMs present flexible and extendable platforms to simulate the behavior of autonomous entities called agents. Wooldridge and Jennings (1995) propose the following definition for an agent: "An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives (Wooldridge and Jennings, 1995, p. 115)". Interacting among multiple interacting agents means intrinsically nonstationary environment for each agent. In this context one agent's behavior might change over time depending on the decisions of its counterparts. Stone and Veloso (2000) propose the following definition for MASs: "MASs are the subfield of AI that aims to provide both principles for construction of complex systems involving multiple agents and mechanisms for coordination of independent agents' behaviors (Stone and Veloso, 2000, p. 345)"

Stone and Veloso (2000) summarize the following useful properties, MASs offer: MASs enable parallelism in computation, since a Multi-agent approach enables the distribution of tasks to different agents. Additionally, MASs offer robustness, because careful distribution of control and responsibilities can lead to tolerance in errors, since the failure of an agent in its task can be compensated by other agents' work. Moreover, another benefit of MASs is their scalability due to the inherently modular design. In other words, it is easier to add new agents to MASs than adding new capabilities to a monolithic system. Furthermore, modularity can also result in simpler programming. Finally, MASs prove to be useful from the perspective of social science as they can aid in the study of intelligence.

In MASs agents must be enough intelligent capable of elaborating decisions on theirs tasks while their knowledge about the system state respectively objectives of other agents is restricted. Agents must instead discover a solution through observation and relying on their own knowledge by *learning*.

1.2.6 *Learning*

Learning and intelligence are closely related to each other. Weiss (2000) defined learning informally as follows: "The acquisition of new knowledge and motor and cognitive skills and the incorporation of the acquired knowledge and skills in future system activities, provided that this acquisition and incorporation is conducted by the system itself and leads to an improvement in its performance (Weiss, 2000, p. 260)". Agent-based computational economists employ various learning models. According to Panait and Luke (2005) there are three main approaches to learning: supervised, unsupervised, and reward-based learning. In supervised learning, an agent deals with the problem of learning the optimal function mapping inputs to outputs by training with a series of input and output pairs. A teacher or supervisor steers the learning progress through providing feedback on the success. Artificial Neural Networks (ANNs) are typically examples of supervised learning. Supervised learning becomes inadequate when the output for a certain input cannot be easily obtained by a supervisor computationally. In unsupervised learning, no feedback is provided. Data mining methods, clustering and discovery are examples of unsupervised learning.

The reward-based learning methods are divided into two subsets: reinforcement learning (RL) and stochastic search methods such as evolutionary algorithms. In RL, agents learn by estimating value functions through delayed rewards. Agents in RL mostly elaborate decisions based on the notion of *dynamic programming* (Bellman, 1957). Dynamic programming solves optimization problems by combining solutions to sub-problems. Each agent solves each sub-problem just once and saves its answer in a table, to avoid the re-computation. In stochastic search methods, agents try to refine their decisions iteratively through testing each possible sequence of actions to find the appropriate one without considering historic states and their corresponding payoffs. Reward based learning models seem the natural choice for the majority of papers in the literature of MASs. A

significant part of it is concerned with RL (Busoniu et al., 2010).Given its relevance, we look into RL in more detail.

1.2.7 *Reinforcement learning*

Sutton and Barto (2005) provide a clear and simple account of the key ideas and algorithms of reinforcement learning. RL describes a non-conscious learning process helping agents to maximize a long-term objective function based on trial and error in a stationary environment. Actions that yield a positive effect will have a higher chance of being chosen again in the future (Sutton and Barto, 2005). The theory of Markov Decision Processes (MDPs) offers a framework for modeling the decision-making procedure by agents in the context of RL. RL uses MDPs for world representation. A MDP (Howard, 1960) is a tuple *(S, A, T, R)*, where S is the set of states, *A* is the set of actions, *T* is a transition function $S \times A \times S \rightarrow [0\ 1]$, and *R* is a reward function $S \times A \rightarrow R$. The transition function defines a probability distribution over the next states as a function of the current state and the agent's action. The reward function defines the reward the agent receives when selecting an action at given state. Solving MDPs consists of finding a policy function μ, $\mu: S \rightarrow A$, which maps states to actions. An optimal policy maximizes the sum of future rewards *r*, discounted by factor γ, over time *t*. The optimal way for agents to learn the optimal policy is learning the optimal *value function* (Sutton and Barto, 2005). The value function v^{μ} is defined *for each state* s as sum of expected discounted rewards r, given the agent follows "some policy" μ starting in that state and following the policy until we achieve a Terminal-state:

$$v^{\mu}(s) = E(\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} | s_t = s, \mu) \tag{1.6}$$

Similarly, a *q-function* is defined as the expected discounted reward given the agent takes a certain action a *in state* s *following policy* μ.

$$q^{\mu}(s,a) = E(\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} | s_t = s, a_t = a, \mu) \tag{1.7}$$

The optimal q-function is defined as $q^*(s,a) = max_\mu q^\mu(s,a)$. It satisfies the Bellman optimality equation:

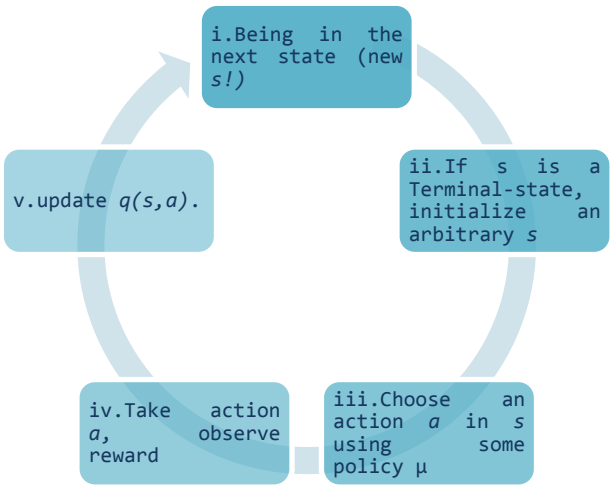
$$q^*(s,a) = \sum_{s' \in S} T(s,a,s')[r(s,a,s') + \gamma max_a q^*(s',a')] \; \forall s \in S \;\; \& a \in A \tag{1.8}$$

The equation (1.8) states that the optimal value of taking a in s is the expected immediate reward from undertaking a plus the expected discounted maximum value attainable from the next state s'. Once q^* values corresponded to actions in each state are available, the optimal policy will be returned in every state by reinforcing the action with the largest optimal q-value.

$$\mu^*(s) \leftarrow arg\ max_a q^*(s,a) \tag{1.9}$$

The optimal policy μ^* of agent in each state would be typically assigning probabilities to actions that obtain higher q-values. A broad range of single and multi-agent RL algorithms are derived from the basic q-learning developed by Watkins (1992).

Figure 1.4: Schematic presentation of q-learning algorithm based on Sutton and Barto (2005).



A q-learning agent maintains the value of each possible action in every state of the environment. These are called *q-values* and are stored in a table. The evaluations of the quality of particular actions at particular states are iteratively improved. The agent, subject to some *error*[7], selects the most favorable action (the action, that gives already the maximum q-value in his current state) *a* in its current state *s*. Then it perceives the consequence of this action in form of the new state of the environment s' and its reward *r*. Through this reward, the agent validates the significance of its last action and updates its q-value. Hence, q-learning turns into an iterative approximation procedure. The agent starts with an arbitrary q-function, observes transitions (s_k, a_k, s_{k+1}, r_{k+1}), and after each transition updates the q-function according to

$$q_{k+1}(s_k, a_k) = q_k(s_k, a_k) + \alpha_k \left[r_{k+1} + \gamma max_{a'} q_k(s_{k+1}, a') - q_k(s_k, a_k) \right]$$

(1.10)

The term within the right bracket is the difference between the current estimate of q-value of (s_k, a_k) and the updated estimate of (s_k, a_k). Parameter setting influences the quality of learning. For example setting factor α_k to 0 means that the q-values are never updated, hence nothing is learned. Setting a high value such as 0.9 means that learning can occur quickly. The discount factor γ describes how an agent will evaluate the rewards, which he gets afterwards. If the discount factor meets or exceeds 1, the q-values may diverge.

1.2.8 *Multi-agent reinforcement learning*

The convergence of RL methods is based on the assumption that the environment is stationary. Learning among multiple strategically interacting agents is far more

[7] For example, in a so called *epsilon greedy* policy an agent chooses a random action with a small probability *epsilon* and with a probability equal to *1- epsilon* decides to take the action, which gives already the maximum q-value in his current state.

demanding task since the choice environment of each agent is now intrinsically non-stationary. Recently there has been growing interest in the literature to extend learning methods to the multi-agent domain. In this case, the state transitions are the result of the joint action of all the agents. Because the rewards of the agents depend on the joint action, their payoffs depend on the joint policy. The question that arises is whether an agent is able to perform effectively without actually taking into consideration the actions of the other participating agents. Experiments have shown that in some cases, agents that learn about the values of joint actions are able to perform better than single learning agents (Claus and Boutilier, 1998). Hence many approaches in the literature aim to extend MDPs to the multi-agent case (Littman, 1994). Despite this, in many multi-agent settings independent single agent learners have shown also good performances (Busoniu et al., 2008). The generalization of the Markov decision process (once multiple agents interact and learn simultaneously) is named "stochastic game". Stochastic games are the extension of MDPs to multiple agents and of static games to multiple states.

1.2.9 *Static games, repeated games and Stage games*

A static (stateless) game is a stochastic game with no state distinction and no dynamics, i.e. $S = \emptyset$. A static game is described by a tuple $< a_1, \dots, a_n, r_1, \dots, r_n >$ with the rewards depending only on the joint actions $r_i: A \rightarrow R$. An important characteristic of a static game is the *Nash equilibrium* of that game. Nash equilibrium is relied on the notion of *best response* of agent *i* to a vector of opponent strategies. A best response strategy σ_i^* is the strategy that obtains the maximum expected payoff given the other players' strategies:

$$E\{r_i|\sigma_1,\dots,\sigma_i,\dots,\sigma_n\} \leq E\{r_i|\sigma_1,\dots,\sigma_i^*,\dots,\sigma_n\} \forall \sigma_i \ . \tag{1.11}$$

The Nash equilibrium describes a joined strategy profile $[\sigma_1^*, \dots, \sigma_n^*]^T$ such that each strategy σ_i^* is a best response to others.

A repeated game is a static game played repeatedly by the same agents. The substantial difference between a static game and a repeated game is that the agents can use the history of the game to learn about the other agents' behavior or

about the reward functions, and make more decent decisions thereafter. A stage game is the static game that arises out of the state *s* of a stochastic game. The reward functions of the stage game in state *s* are the q-functions of the players projected on the joint action space, when the state is fixed at *s* (Busoniu et al., 2010).

1.2.10 *Practical multi-agent learning*

A comprehensive taxonomy of multi-agent learning algorithms in general can be found in the work of Busoniu et al. (2010). Multi-agent reinforcement algorithms are predominantly applicable to small problems only, like small stochastic games and small grid worlds. Scalability is a central concern of MARL. By increasing the interaction domain tabular storage of q-functions for agents becomes economically infeasible i.e. impractical. Hence, there is continuous research towards the development of robust agents for large-scale, complex, open, dynamic and unpredictable environments (Busoniu et al., 2010). In the field of computational economics, some researchers prefer to develop interaction models on the basis of the individual based psychological RL methods. For example, Nicolaisen et al. (2001) use routine based Roth-Erev model (Roth and Erev, 1995) to investigate market power and efficiency outcomes for a short-run, wholesale electricity market with double- auction pricing and with buyers and sellers who continually update their price offers on the basis of past profit experiences. In the oligopoly and oligopsony learning literature, especially in spatial competition of agricultural markets, the most widely spread practical methods to analyse the firms' behavior are evolutionary methods used e.g. in Graubner et al. (2011a and 2011b).

1.3 **Contributions, key results and limitations**

Each of three main chapters of this thesis takes us through some basic problem for designing autonomous learning agents in dynamic and spatial ABMs.

Chapter 2 of the thesis may be seen as the motivator study of the thesis. It models an interactive environment comprising possible *asymmetric* pricing policies between firms, as well as policies involving *freight absorption*. It serves especially as a key theory for understanding how collusive behavior might have

become a prevailing pattern in many agricultural markets. For example, the studies of Huber (2007 and 2009) show that in almost half of the area of Germany, more than one dairy processor agents operate in certain regions. Cooperation between major processors can be inferred also from numerous empirical and theoretical studies e.g. the milk inquiry of the German cartel authority (section 1.2). Indeed milk processors coordinate their prices based on a regional average milk price of other processors (Bundeskartellamt, 2009). Our study offers insight into the price-matching rationality of processor firms observed in raw milk markets. Herewith we constitute a simplified spatial configuration for a procurement market by considering two milk processors at the end of a line containing supplier farms. In the classic economic models it is typically assumed that sellers (e.g. here milk suppliers in our model) in procurement markets deliver the raw product to processor firms and receive the mill price at the buyer's factory gate (FOB pricing). Here we presume that farmers receive the same price at their *farm gate* irrespective of their location relative to the processor's production plant (UD pricing). Empirical studies frequently observe the application of UD pricing rules for real. A major problem encountered in the current literature on spatial competition with uniform delivered pricing policies however, is the nonexistence of pure-strategy Nash equilibria in competitive models due to discontinuous best response function of players (Dasgupta and Maskin, 1986; Beckmann, 1973; Schuler and Hobbs, 1982). In such cases, cyclic price wars take the place of Nash equilibrium. System characteristics such as limitless price wars have been made clear previously in the classic models of pricing in Shubik (1980) as well as in spatial competition models (Schuler and Hobbs, 1982). The lesson we learn from our study is not only how spatial characteristics of markets lead "myopic rational" agents to get bogged down in interminable non-cooperative price disputations, but also how imposing a minimal degree of awareness regarding competitor's possible reactions in the agents' pricing decision, might invite decent agents to revise their primary malicious based decision. We show how *foresight* based competition might force players' policy to deviate from cyclic price patterns and move towards cooperation. One party not only thinks "What happens to my payoff if I overbid?" but also "Will I be overbid by the opponent player after I have already overbid?" and "what is going to be my next reaction after my opponent's reaction?" The ability of real human agents to reflect on their behavior based on

subsequent consequences of their myopic decisions will distinguish them from other non-intelligent decision systems. We suggest that one substantial *institutional* feature of milk market namely *contractual relationships* dominating the market-relationships might expedite the foresight-based decision. In contrast to spot markets where market participants can act like myopic responder agents, in long-term contractual relationships one firm will necessarily try to interfere in market of its competitor after he learns that its opponent has achieved new arrangements with its suppliers. The intervening party might try to spoil the business plan of the opponent. The opponent might try to convince its contracted suppliers not to switch. We note that without considering such institutional aspects (*contract*- and *auction*-based interactions between market participants), many empirical and agent-based investigations would have overlooked some strong implications of contract or auction regimes.[8] Though the full implications of institutional aspects of markets, e.g. contractual relationships between market participants, are rarely investigated in prior studies of spatial competition in agricultural markets, their influence upon competition and welfare might be pervasive.

An additional relevant contribution of chapter 2 is the correlation between firm locations and collusive behavior in light of aforementioned institutional factors. The correlation problem has been studied well analytically in the literature of horizontal product differentiation too.[9] Results from a small body of literature indicate that a smaller firm dispersion is more likely to sustain tacit collusions. The majority of studies suggest that the relationship between higher degrees of product differentiation and collusive behavior is robust. In our model, a larger spatial dispersion of firms might grant a greater *outside* utility and consequently less competition in the market. However, our model counts for the effect of institutional factors as a major caveat to understanding real market interactions. It proposes that advantages of strategic thinking for players within the system are

[8] The relevance of contract theory is addressed mostly in the works of Oliver Hart, the Nobel Prize winner of economics in 2016. See e.g. Hart (2009).

[9] Differential factor for different products can be translated into how close consumers (or here farmers) are to each firm stand.

larger, if the importance-of-space is smaller. In fact, with lower factor importance-of-space, competitor agents are supposed to learn more from each other by building price conjectures about the opponent's compensatory reactions. In such interactive market structures, it is more probable that decent competitors recognize the value of cooperation.[10]

Meanwhile some theoretical hurdles limit the application of our model to reality. A crucial limitation to establish the proposed equilibria in chapter 2 in learning domains might be that all agents need to recognize the same world. Agents are able to convey the shared understanding that unilaterally switching their pricing policy based on malicious decisions would be credibly prohibitive. The need for coordination of self-interested agents in our model is resulting from the dependence of any agent's action on foreseeable actions taken by the counterparts. However, in many real learning domains, agents might incorporate some wrong beliefs toward specific compensatory actions taken by other agents. This might restrict application of our theoretical model in chapter 2 to real learning domains. One might note that, achieving such cooperative rationality levels might be feasible through a *structured dialog* between agents during the course of interaction. In fact, real human businesses do use *high-level* communication forms in a wide variety of their activities. Such learning procedures based on high-level communicative interactions like verbal negotiation and mutual explanation is studied less in the literature.[11]

In chapter 3, we propose a new decision making algorithm, called H-PHC, which enables multiple agents to learn in rich strategic decision spaces without explicitly modelling beliefs about the interaction environment and other participating

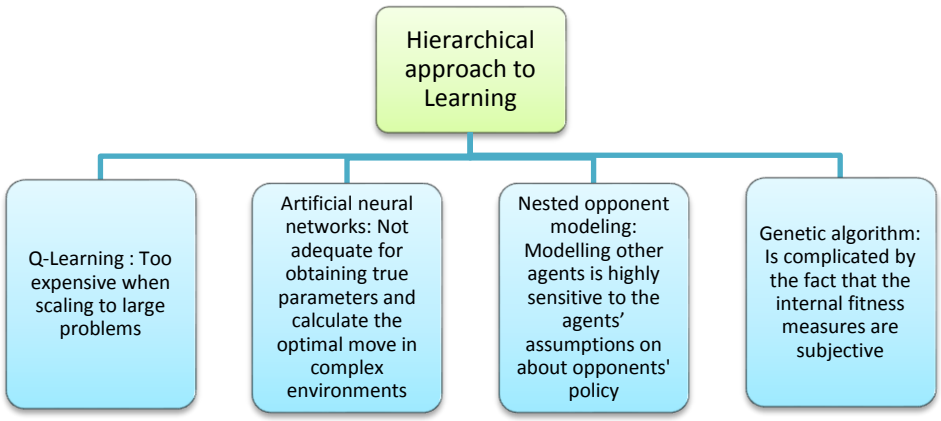
[10] We will draw similar inferences regarding the interrelation between importance-of-space and collusive behavior in Chapter 4.

[11] The idea of complex communication can be pursued in the literature of AI. For example in Sian (2005) multiple agents, each of them trading agricultural commodities in particular area (Kenya, Brazil and India) aim at cooperatively create generalized descriptions of how prices of tea, coffee and cocoa change due to various classes of events (flood, frost and drought) through explicit communication protocols.

agents. One intrinsic property of rational human decisions guided us to develop the algorithm: humans are used to hierarchically elaborate their decisions rather than to undertake an exhaustive enumerative search through the whole decision space. This avoids that agents change their decisions in an arbitrary manner and will reduce the dimensionality of interactions in real life procedures. For example, once an agent has chosen the spatial location for its processing plant, the computational efforts regarding optimal pricing interactions with neighbouring processors is considerably diminished (compare to before the decision). We suggest that such hierarchical rationality of agents is a *sufficient condition* to learn in rich strategic spaces with modest computational effort. It is significant that the H-PHC algorithm does not assume agents who know the underlying game or have prior knowledge regarding the corresponding Nash Equilibrium.[12] The only feedback one agent might need is its own local reward as the agent does not explicitly model other agents' actions or rewards. Learning buyer (milk processor) agents applying H-PHC algorithms in our study show to fair efficiently reaching the Nash equilibria. The advantage of our approach in comparison with alternative methods is demonstrated in Figure 1.5.

[12] Having prior knowledge regarding firms' production function or best response functions is presumed as a pre-requirement of the learning procedure, e.g. through ANNs in Barra and Saracenob (2005).

Figure 1.5: Advantages of hierarchical approach to learning over alternative methods in the learning literature of pricing and quantity games.



Our experiments give evidence that our algorithm succeeds to act as a cognitive motor encouraging non-cooperative coordination not only in various spatial games but also in variety of non-spatial markets.

Despite this accomplishment, further improvement of H-PHC as a rational algorithm to guide agents' decisions in non-trivial dynamic markets seems necessary. For example, one problem to be surmounted is the non-existence of Nash equilibrium in the presence of freight absorption policies (e.g. UD pricing). In such environments - without communication or knowing more about each other's objectives - rational competing agents must continually adapt to each other again and again, never stabilizing at an established behavior. A rational learning algorithm is supposed to reflect the price instability phenomenon. By means of an example we show that introducing some extra decision factor into the algorithm (Meta-Decision), the frequent non-stationary cycling price procedures can be made possible. However, designing H-PHC agents able to shift their decisions endogenously whenever agents' policy must be revised (through rationality criterion) is not completely performed. This is still an open issue to be improved. Indeed the remaining concern for H-PHC is how to learn rational and how fast. A dynamic market environment can comprise variety of endogenous shocks e.g. endogenous changing of farms' supply behavior or evolving reaction functions of some firms and etc. Relaxation of quota regulations in agricultural markets in the EU is an example triggering such environmental shocks. Designing

highly sensitive rational agents equipped with some well-designed Meta-Decision module would be highly interesting. One remedy to implementation of such Meta-Decision mechanism could be learned from the well-known Wolf-PHC (Bowling and Veloso, 2002) algorithm, in which the designer *imposes* on agents to learn (new policies) slower when they are winning and faster when they are losing.

Lastly one might note that the rationality levels – based on myopic rationality of humans – such that we spotlight in chapter 3- are *primitive* to describe the real learning processes of real economic agents. As revealed in chapter 2, we need to incorporate some higher levels of rationalities in order to capture empirical evidence regarding collusive behavior of real agents. Chapter 4 elaborates on this issue.

In chapter 4, we investigate a learning based analysis of competition in a duopsonistic milk processor market in two-dimensional space. A buyer agent can offer each seller varying prices depending on the distance of that farm to the processor's location. The analysis is performed in two scenarios reflecting two opposite poles of our understanding about learning aptitude of processors, *low-coordination* and *high-coordination*. First, we create agents who are born with some previously acquired knowledge about observing the state of the world and using the utility maximization rule of best myopic response to boost their profit. We name theses agents *A-level* agents. We design a dynamic sequential interactive game by simulating their behavior. The process in which agents will take turns setting prices that are the best myopic response to the opponent is analogous to the process Cournot studied and is expected to have the same long-run property as the simultaneous move adjustment process (Fudenberg and Lewin, 1998, p.11). We propose that the price competition of A-level agents will always be describing a self-enforcing pattern regarding its direction. Principally the system moves always toward some unique cyclic basin of attractions of pricing contest (Terminal-states).[13]

[13] Generally, we presume that Terminal-states represent close to Nash equilibrium strategies of the play, since they describe a local target in which the strategies of players will be trapped.

In the second stage, we investigate some type of higher rationality levels in decision procedure of agents. We incorporate the fact that real economic agents are *dynamic programmers* i.e. they can store predicted knowledge linked with forthcoming phases of bilateral price interactions and decide upon this knowledge while deliberating in advanced stages of policy setting. Basically, recursive modeling agents have been adopted from the *minimax* heuristic search in games like chess (Carmel and Markovitch, 1996). We define *B-level* and *C-level* agents as the agents who (compared to A-levels) act more strategically. These agents incorporate conjectures on upcoming reactions of their counterparts. Expected behavior of the other agent and the agent's own optimal policy in the system in each forthcoming stage of the world get stored. Each time agents face the same sub-stage of the game they search for the previously known information regarding the upcoming stages of interaction. The agents then forecast system-wide consequences of their pricing behavior for their overall payoff from each current state of the world onward. This distinguishes our study from most theoretical studies, which have investigated the problem of policy making in spatial markets through understanding the equilibrium of interactions in *static* settings.

On the basis of the aforementioned rationality levels, we draw inferences on pricing behavior of agents. The results of our study differ to some extent from the prior literature. Whereas predecessor papers often have attributed the degree of freight absorption by processing firms merely to the spatial structure of markets, we suggest that in addition the pricing behavior of agricultural processors also depends on their ability to learn from each other. Furthermore, our simulation outcomes reveal that in a world where agents learn from each other, possible permissible collusive pricing by agents might compromise a variety of freight absorption as well as mill prices. The more the factor importance-of-space increases, the more permissible collusive pricing actions will be assembled around the well-known optimal price discrimination (OD). This is because the larger the distance of firms in space the less fruitful are efforts regarding learning about the opponent compared to applying the best myopic response.

Our simulation model shows that processor agents are better off in circumstances when they learn from each other rather than situations where they would not learn from each other. We like to give particular emphasis to the fact that Nash equilibrium predictions are not necessarily the efficient outcomes of markets, but

simply the one that will result when each player in the system is individually pursuing his own optimal myopic utility function (A-level perception) without performing learning. Like Shoham et al. (2004) we believe that learning procedures that specifically target Nash equilibria are troubling to be prescriptive:

"*Nash equilibrium at the best identifies conditions under which learning can or should stop but it does not purport to say anything prior to that* (Shoham et al., 2004, p.3)".

In summary, it is worth noting that *Learning* in spatial systems requires algorithms that are scalable to a large number of agents and can be implemented with minimal knowledge about the actions of other agents. Most proposed multi-agent learning algorithms in the literature fail one or both of these criteria. In addition, a key issue is to distinguish between recent works in computational economics and MASs. Although AI science and computational economics are related fields of research they seem to have progressed along separate paths. We believe significant progress in the field of ABMs can be achieved by more intensive cross-fertilization between the fields of machine learning and computational economics. This thesis proposes initial viewpoints towards how to involve adaptation sense of intelligent agents in agricultural agent based models.

1.4 References

Appelbaum, E. (1982). The Estimation of the Degree of Oligopoly Power. Journal of Econometrics 19, pp. 287-299.

Azzam, A. M. and Pagoulatos, E. (1990). Testing Oligopolistic and Oligopsonistic Behavior: An Application to the US Meat Packing Industry. Journal of Agricultural economics 41, pp. 362-370.

Babes, M., Cote, E. M. and Littman, M. L. (2008). Social Reward Shaping in the Prisoner's Dilemma. Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008).

Barr, J. and Saraceno, F. (2005). Cournot competition, organization and learning. Journal of Economic Dynamics & Control 29, pp. 277–295.

Bellman, R. (1957). Dynamic Programming. Princeton, New Jersey: Princeton University Press.

Bowling, M. and Veloso, M. (2001). Rational and convergent learning in stochastic games. In: Proceedings 17th International Conference on Artificial Intelligence (IJCAI-01), pp. 1021-1026. San Francisco.

Brenner, T. (2005). Agent Learning Representation –Advice in Modelling Economic Learning. Germany: Max Planck Institute for Research into Economic Systems.

Britz, W. (2013). A flexible framework for ABMs to simulate spatially explicit structural change in agriculture, Methodological and technical documentation, Version 1.0. Institute for Food and Resource Economics, University of Bonn.

Bundeskartellamt. (2012). Inquiry Milk sector. Bonn, Germany: Federal cartel office.

Busoniu, L., Babuska, R. and De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 38 (2), pp. 156-172.

Busoniu, L., Babuska, R. and De Schutter, B. (2010). Multi-agent reinforcement learning: An overview. In: D. Srinivasan, L. Jain (eds.) Innovations in MASs and Applications - 1, Studies in Computational Intelligence, vol. 310, pp. 183–221. Berlin Heidelberg: Springer.

Claus, C. and Boutilier, C. (1998). The dynamics of reinforcement learning in cooperative multiagent systems. Proceedings of the Fifteenth National Conference on Artificial Intelligence, pp. 746–752.

Fudenberg, D. and Levine, D. K. (1998). The theory of learning in games. Cambridge, MIT Press.

Graubner, M., Balmann, A. and Sexton, R.J. (2011b). Spatial price discrimination in agricultural product procurement markets: A computational economics approach. American Journal of Agricultural Economics, Vol. 93(4), pp. 949-967.

Graubner, M., Koller, I., Salhofer, K. and Balmann, A. (2011a). Cooperative versus Non-cooperative Spatial Competition for Milk. European Review of Agricultural Economics. European Review of Agricultural Economics, Vol. 38(1), pp. 99-118.

Grimm, V. and Railsback, S. F. (2005). Individual-based Modeling and Ecology. Princeton University Press.

Grossman, S. and Hart, O. (1986). The costs and benefits of ownership: a theory of vertical and lateral integration Journal of Political Economy 94(4), pp. 691-719.

Hart, O. 2009. “Hold-Up, Asset Ownership, and Reference Points.” Quarterly Journal of Economics 124 (1), pp. 267–300.

Hellberg-Bahr, A., Pfeuffer, M., Steffen, N., Spiller, A. and Brümmer,B. (2010). Preisbildungssysteme in der Milchwirtschaft, Ein Überblick über die Supply chain Milch. Göttingen: Department für Agrarökonomie und Rurale Entwicklung, Universität Göttingen.

Howard, R. (1960). Dynamic programming and Markov process. MIT Press.

Kirman, A. (2011). Learning in ABMs. Eastern Economic Journal 37 (1), pp.20–27.

Klein, B., Crawford, R. and Alchian, A. (1978). Vertical integration, appropriable rents,and the competitive contracting process. 21, pp. 297–326.

Koller, I. (2012). Spatial pricing and competition:A theoretical and empirical analysis of the German raw milk market. PhD Thesis. Technischen Universität München.

Littman, M. L. (1994). Markov games as a framework for Multi-agent reinforcement learning. In Proceedings of the Eleventh International Conference on Machine Learning, pp. 157-163. Morgan Kaufmann.

Nicolaisen, J., Petrov, V. and Tesfatsion, L.,”Market Power and Efficiency in a Computational Electricity Market with Discriminatory Double-Auction Pricing”. IEEE Transactions on Evolutionary Computing 5, 5 (October 2001), pp. 504–523.

Panait, L. and Luke, S. (2005). Cooperative Multi-agent Learning: The State of the Art. Autonomous Agents and MASs Volume 11 Issue 3, pp. 387 - 434.

Wooldridge, M. and Jennings, N. (1995). Intelligent Agents: Theory and Practice. The Knowledge Engineering Review, vol. 10, no. 2, pp. 115–152.

Osborne, M.J. and Rubinstein, A. (1990). Bargaining and Markets. New York: ACADEMIC PRESS, INC.

Perekhozhuk, O., Hockmann, H., Fert, I. and Zoltán Bakucs, L. (2013). Identification of Market Power in the Hungarian Dairy Industry: A Plant-Level Analysis. Journal of Agricultural & Food Industrial Organization 11(1), pp. 1-13.

Sian, S. (2005). Extending learning to multiple agents: Issues and a model for multi-agent machine learning. Machine Learning — EWSL-91, pp 440-456.

Stone, P. and Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. Autonomous Robots, vol. 8, no. 3, pp. 345–383.

Rogers, R.T. and Sexton, R.J. (1994). Assessing the Importance of Oligopsony Power in Agricultural Markets. American Journal of Agricultural Economics, Vol. 76, No. 5, pp. 1143-1150.

Tribl, C. and Salhofer, K.(2013). Marktmacht und räumlicher Wettbewerb entlang der Wertschöpfungskette von Milch. Wien: BUNDESANSTALT für Agrarwirtschaft.

Roth, A.E. and Erev, I. (1995): "Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Run," Games and Economic Behavior, 6, pp. 164-212.

Sexton, R. J. (1990). Imperfect Competition in Agricultural Markets and the Role of Cooperatives. American Journal of Agricultural Economics, Vol.72 (3), pp. 709-720.

Sexton, R. J. (2012). American Journal of Agricultural Economics, Volume 95, Issue 2, 1 January 2013, pp. 209–219.

Shoham, Y., Powers, R. and Grenager, T. (2004). On the agenda(s) of research on Multi-agent learning. In Proceedings of Artificial Multiagent Learning. Papers from the 2004 AAAI Fall Symposium. Technical Report FS-04-02.

Sutton, R. S. and Barto, A. G. (2005). Reinforcement Learning: An Introduction. Cambridge, Massachusetts: The MIT Press.

Watkins, C.I.C.H. and Dayan, P. (1992). Q-learning. Machine Learning, 8, pp. 279–292.

Watkins, C.I.C.H. (1989). Learning from delayed rewards. PhD thesis, University of Cambridge, Cambridge.

Weiss, G. (2000). Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press.

Wieck, C. and Mosnier, C. (2011). Determinants of the spatial structure of regional production system. European Association of Agricultural Economics Conference. Zurich.

Chapter 2
Outside Option and cooperative behavior of learning agents in spatial markets

Abstract. The recent literature on spatial competition shows empirical evidence for cooperative pricing in agricultural markets especially when it comes to the spatial pricing in raw milk markets. While the phenomenon of collusive behavior underpinning such outcomes is often investigated, static models predict that Nash equilibria in pure strategy cannot exist due to the discontinuous nature of players' best response function. We design a dynamic model of spatial competition and investigate how a coordinative pure strategy Nash equilibrium might arise in a broad range of market structures if agents are actors with foresight, who learn from each other's upcoming actions. A higher share of raw milk committed via long-term supply contracts, i.e. not freely available on the spot market, may reinforce the equilibrium.

Keywords: Oligopsony, Agricultural spatial markets, learning

JEL classification codes: **D43, L13, L40, Q10, C63**

2.1 Introduction

Food processing has experienced a substantial increase in market concentration (Sexton, 2012). A number of studies emphasize that agricultural markets are oligopsonies. Many spatially dispersed suppliers face relatively few processors of raw products with relevant transport costs (Sexton, 1990 and 2012). The raw milk market is an example of an agricultural procurement market with high transactions costs and thus likely strong market power. High repercussions from

volatilities of supply-demand in the downstream market influence the prices at farm level. Farmers use specific assets to produce raw milk which is a perishable good. They cannot employ their farms, cows and machinery for other purposes during periods of low prices (United States accountability office, 2004, p.97). Indeed, the perishability of raw milk can put farmers in against-the-clock situations which enhance the counterparts' bargaining powers i.e. dairy processors (European Commission, 2013, p.36). The number of dairy processors and the number of operating sites in Germany show a long term declining trend, a pattern that is repeated in most other EU countries (Boysen and Schröder, 2006). Several studies have recently investigated potential collusive behavior on the part of the milk processors (Graubner et al., 2011a; Huber, 2009; Bundeskartellamt 2009; Huber, 2007 a; Huck et al., 2006; Alvarez et al., 2000). Huber (2007 and 2009) show that more than one processor operates in almost half of the area of Germany, and regions exclusively served by one processor generally have a low density of milk (Huber, 2009, p.36). The German cartel authority found evidence suggesting potential collusion when they showed that milk processors set their prices based on regional average prices of other processors (Bundeskartellamt, 2009).

Despite the substantial evidence of cooperation among dairy firms in practice, theoretical justification of cooperative and non-cooperative equilibriums in suchlike markets are not well sustained. Theoretical investigations of strategic interaction of firms in spatial agricultural markets are restricted by using a predefined set of specific behavioral assumptions regarding price setting policies of oligopoly firms, the so called conjectural approaches (Capozza and Van Order, 1978).[14] For example Graubner et al. (2011a) test whether price transmission from

[14] A conjecture is defined as the reaction of a firm to a change in a competitor's price. In spatial competition, several conjectures are distinguished. Under the Hotelling-Smithies (HS) conjecture (Hotelling, 1929), a processor takes the competitor's price as given when deciding on the own pricing policy. The Lösch conjecture (Lösch, 1954) presumes full coordination of the firms regarding their pricing policy: processors cooperate to maximize joint profits implying a cartel solution. According to the Greenhut-Ohta conjecture (Greenhut and Ohta., 1972), each firm assumes that its competitor reacts to an own price change with the same price change in the opposite

the producer to the wholesale price in the German raw milk market is consistent with certain conjecture. They suggest that the payoff matrix defined by cooperative and non-cooperative pricing policies has the structure of a prisoner's dilemma. According to their study, the observed low price transmission in Germany is in line with the price-matching conjecture although, in a static Nash equilibrium, both processors should deviate from cooperative behavior and opt for non-cooperative pricing, i.e. Hotelling-Smithies conjecture.

On the other hand the majority of studies assume the application of uniform delivered (UD) pricing by dairy processors. Indeed UD pricing is a basic constituent of several empirical and theoretical contributions investigating the behavior of processors (Graubner et al., 2011b; Alvarez et al., 2000; Huck et al., 2006; Tribl, 2012).

A major inconsistency arises from simultaneously incorporating freight absorption (e.g. UD pricing) and cooperative conjectures (e.g. PM conjecture) in the interaction of firms is the Nonexistence of pure-strategy Nash equilibria in static competitive models comprising freight absorption policies due to discontinuous best response functions (Dasgupta and Maskin, 1986; Beckmann, 1973; Schuler and Hobbs, 1982). In the non-existence of Nash equilibrium it is expected that (in line with the rational behavior of agents observing the action of opponents) unsteady price battles occur. This fact might cause that sustaining cooperative interaction among processors would not be feasible. Pathological cyclic price behaviors had been figured out previously in the classic models of pricing in (Shubik, 1980) and are also discussed for spatial competition models (Schuler and Hobbs, 1982).

---

direction. Lastly the price-matching (PM) conjecture (Gronberg and Meyer, 1981; Alvarez et al., 2000) proposes a mutually agreed price commitment of market participants as the core assumption under spatial competition.

We argue in this paper that dynamic pricing policies by processor agents help them to escape coordination failure. We focus on understanding how dynamic forces in spatial market environments may create stronger incentives for cooperation than those in a static game. We show that rational cooperation occurs if agents are endowed with some minimum level of prediction capability with respect to the upcoming stages of the game.

We additionally propose that contractual relationships in the milk market expedite the described foresight-based decision making. In contrast to spot markets where market participants are more likely to act like myopic responders, firm in long-term contractual relationships is more likely to learn about the market planning of its competitor through its new arrangements with its suppliers, and would be likely to strategically respond. For example, firm might try to convince its contracted suppliers not to switch in response to a new contract offered by a competitor. The research of German cartel office (Bundeskartellamt, 2009, p.73) upholds the fact that a high share of milk is committed in the long term via supply contracts, i.e. it is not freely available on the spot market. Indeed there might be several reasons why such institutional aspects are established practices in milk markets. From dairy farms' perspective without contractual or property rights arrangements making sunk costs of milk production is not compatible with the ex-post opportunistic behavior of dairy processors (see e.g. Hart, 1995). From this point of view the crucial assumption in our study comprises that long term contracts might encompass more closed relationships between processor agents and their contracted suppliers and may open up more compensatory reactions for processors by reacting to the potential of their suppliers to sell to another firm.

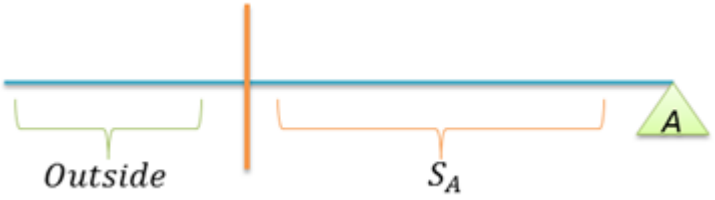
The remainder of the paper is organized as follows: In the next section, we develop a basic dynamic scenario of spatial competition in the context of a duopsony in a hypothetical procurement market. The focus is on the main characteristics of market system behavior in a spot market and how that may change if competitors are able to make contractual compensations by means of compensatory supply contracts. In section 2.3, we show how processors' subjective beliefs about the competitor's ability to make contractual compensations may lead to coordinative behavior in the market. In the last section, we discuss the implications of the results in markets with inelastic supply and in the cased of disparity between firms's down-stream market attributes.

2.2 **The model**

We consider two risk neutral processors operating in a three period game. In *t=0*, dairy processor *A* is a monopsonist who has contracts with dairy farms that evenly distributed with density *D = 1* on a one-dimensional area *S* (line market). Each farm produces a homogenous raw product according to the simple supply function $q = u^{\varepsilon}$ where ε is the price elasticity of supply for a single farmer and *u* is the UD price paid by the dairy. In *t=1* we create dairy *B* just at the market boundary line of *A. B* decides then whether to enter the spatial competition with firm *A.* In *t=2, A* decides whether to react to the entry decision of *B.* If the game is not terminated at *t=2,* then both parties enter some spatial price competition in *t=3*.Neither dairies nor farmers have capacity constraints. Under the assumption of UD pricing, processing firms are responsible for transportation costs (*t*) per unit times distance (*r*) so that each farmer receives the same price. Setting production costs to zero in the downstream market, processors *A* and *B* receive prices P_A and P_B, respectively. Processors maximize profit from buying, transporting, and processing the raw product and selling their product to the final market.

2.2.1 *Basic scenario*

Figure 2.1: Line market and outside area for the monopsonist *A*.



Let's assume that S_A and u_A are the monopsony area and milk price of dairy *A*, respectively.

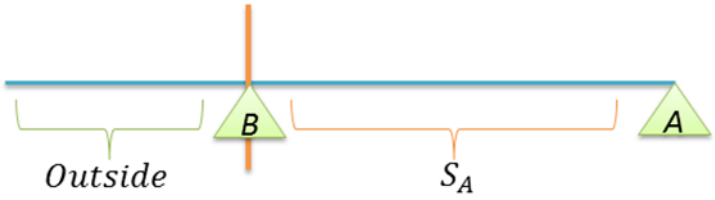
Solving

$$\frac{\partial}{\partial i}\int_0^s u^{\varepsilon}\left(P_A - u - tr\right)dr = 0, i \in \{s,u\} \tag{2.1}$$

This returns optimal area and price of *A* as $S_A^{opt} = 2P_A/(2+\varepsilon)t$ and $u_A^{opt} = 2P_A/(2+\varepsilon)$. One can easily see from derived optimal values how increasing the parameter price elasticity of supply ε will reinforce the Monopsony area and price towards zero respectively P_A.

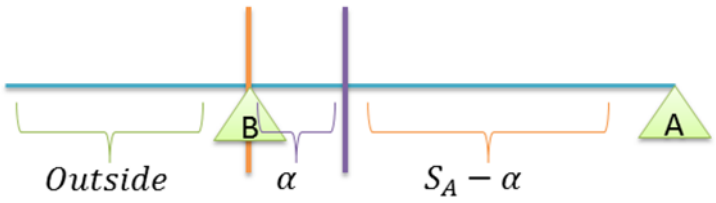
The outside area (see Figure 2.1) describes the area in which milk procurement is not economical for A. Let us first assume that contracts cannot be cancelled by the processor in the short run, but each farm has the right to terminate the contract at any time (temporary assumption 1)[15]. In *t=1* we create dairy *B* just at the market boundary line of *A* (temporary assumption 2).

Figure 2.2: Location of potential entrant *B*.



Locating *B* this way allows *A* to still act as the monopsonist in its area S_A but requires agent *B* to make an inevitable decision regarding entering spatial competition with *A*. Assume both firms are commonly informed about following scenarios: in *t=1*, *B* decides to enter the spatial competition with firm *A* invading a presumed fixed small portion Alpha of *A*'s territory (*S-Alpha* > *Outside* > *Alpha*) hoping to procure the input in this area (fixed disputed area is temporary assumption 3). Alternatively, *B* decides to be just outside of *A*'s market area in *t=1* and exerts another monopsony power just in area Outside.

Figure 2.3: Dairy *B* may invade area *Alpha* in *t=1*.



In *t=2*, *A* reacts to the entry decision of *B*. In case of entry, *A* may give up the invaded area *Alpha* and exercises monopsony market power just in the residual

[15] We use temporary assumptions 1-3 to graphically introduce the Outside Option of an agent and will relax them in our subsequent mathematic setting in section 2.3.

area *S-Alpha* and the game terminates. Alternatively, *A* decides to enter spatial competition in *t=2* hoping to regain possession of the supply from those farmers in *Alpha*. In fact if the game is not terminated at *t=2*, then both parties are spatial price competitors in *t=3*.

The derivation of a sub-game perfect equilibrium of this game depends on what will happen at stage 3 (backward induction). Specifically, each firm's behavior hinges upon the prediction of the opponent's price setting behavior in *t=3* (i.e. their price conjecture). The outcome of dynamic interactions between *A* and *B* in *t=3* will be the focus of our discussion in the following.

2.2.2 *Basic propositions*

Although price competition in *t=3* exists in area *Alpha*, processors are allowed to exert a primary price scheme in their 'backyard' next to their own factory away from the opponent, i.e. the areas Outside for *B* and *S-Alpha* for *A*. We name these two distinct areas the Outside Options of the firms in our basic scenario. However, we relax the fixed Outside Option assumption analytically and will precisely define Outside Option of firms further below. Based on the basic scenario we make the following propositions:

Proposition 1: Each agent knows the maximum price set by each agent in price competition of t=3.

Proposition 2: Undesirable system characteristics such as endless price wars can occur in an economy of myopic agents trading in a hypothetical spot market.

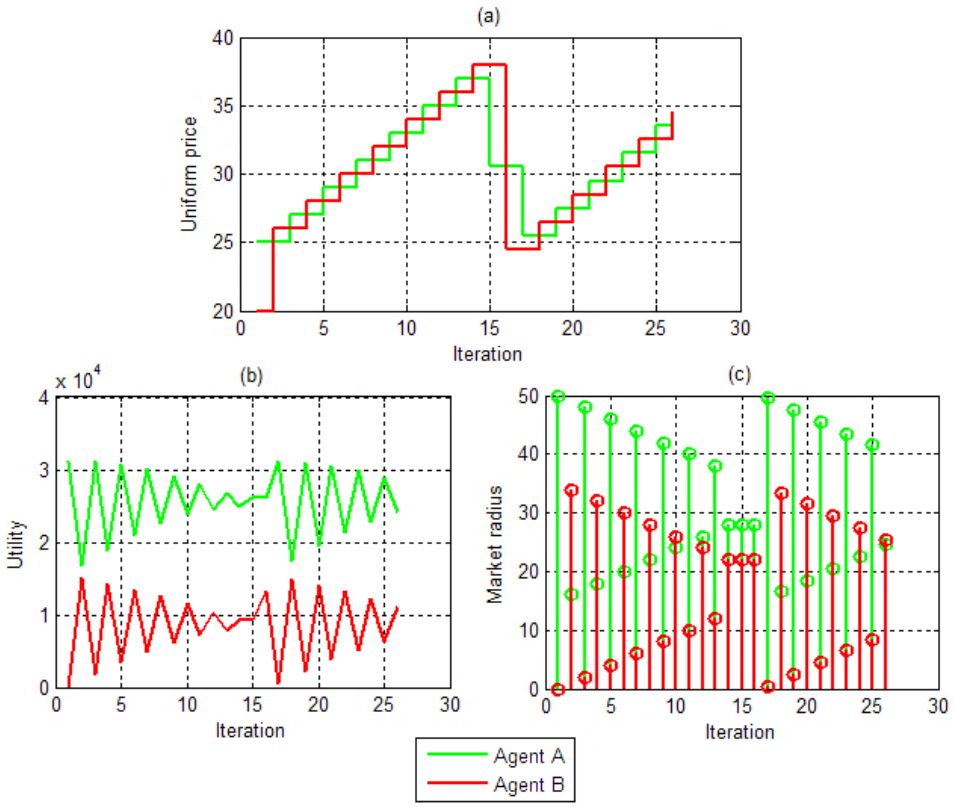
Proposition 3: Introducing anticipation to the agents' pricing policy moves cyclic price wars towards cooperation.

The first proposition is based on the logical reasoning of both players in line with the notion of forward induction (Van Damme, 1989). Both agents might be aware that the maximum price set by each agent in *t=3* is bounded by the net utility of this agent's Outside Options in aforementioned backyard areas. Indeed, the utility associated with retreating to the outside area acts as a signal for the competitor in the spatial price competition. The logic of forward induction implies that the rational players may reason from the very beginning of the game trees as well as they are supposed to be able to argue from its end (backward induction). Being in

the game together with *B* at *t=3*, *A* knows from *t=1* of the game that by turning to the new UD pricing strategy, *B* has foregone his Outside Option characterized by its monopsony pricing scheme. *A* may conclude that *B* is ready to set at most a price bounded by the net profit of *B*'s Outside Option (utility of monopsony *B* in Outside). The same awareness exists for *B* about the refusal of *A* to set a price leading to a utility below the one obtained by exerting monopsony power in *S-Alpha*. This knowledge of agents regarding maximum aspiration levels of each other thorough setting higher prices can help us to understand the relevance of proposition 2. Let's imagine a series of distinct sequential interaction stages at *t=3* of the game. The play begins with the action of player *B* after deciding to enter. In each stage of the game just one of agents decides upon price given the price of the opponent. In order to illustrate the whole pricing procedure we would like to simulate the described statement of proposition 2. For this we relax the temporary assumption 3 (regarding the fixed disputation area) in Figure 2.3. Exactly derived utility functions of firm *A* ($\Pi_A^{u^A}$) and firm *B* ($\Pi_B^{u^B}$) in section 2.3 are the basis of the agents' decision presented in Figure 2.4. One can see how iteratively setting the best policy for each agent starting from any arbitrary state will lead to the cyclic price setting mechanism suggested in proposition 2.

Figure 2.4: Cyclic price setting behavior illustrating proposition 2.

Note: Downstream market prices of *A* = 75 and *B* = 60. Transport cost and price elasticity of supply are set to 1. Distance between firms *R=50*. Overbidding margin = 1.



We can start from the initial point of *B*'s creation where the firm *A* buys the raw product for a price equal to its monopsony price (i.e. u_A^{opt}= 25 in Figure 2.4). The follower firm *B* may overbid the price of *A* by setting a price marginally higher than *A* (e.g. assume an overbidding margin =1). As a response to *B*'s decision, *A* may also overbid the price set by *B* by a margin.[16] Both firms may continue to bid up the price. Overbidding will stop when further overbidding no longer allows

[16] The process in which agents will take turns setting prices that are the best myopic response to the opponent is analogous to the process the Cournot studied and is expected to have the same log run property as the simultaneous move adjustment process (Fudenberg and Lewin, 1998, p.11).

firm *B* to steal firm *A*'s market area. The stopping point represents the point at which overbidding ceases to be useful for *B* (see proposition 1). When firm *B* reaches this point, it decreases the price to the value making *B* again to the residual claimant of the market in area outside. *A* now also lowers its price to some point marginally above *B*'s price since it doesn't make sense for *A* to set a high price without *B* following. However the overbidding cycle is reinitiated as *B* observes the new situation triggered by *A* and tries to overbid again.

Classic models of price wars, including those introduced by Cournot and Bertrand (Tirole, 1988) have the feature that prices are driven down to a minimum value (e.g. the marginal cost in Bertrand's model). However, limit cycle price wars had been observed primarily in the literature in a simple model introduced by Edgeworth (Shubik, 1980) and is discussed in spatial competition models (Schuler and Hobbs, 1982), in MASs (Tesauro and Kephart, 1998) and in evolutionary learning algorithms (Luke and Wiegand, 2002).

As we have seen, market price oscillations occur where processors operate like myopic agents in a spot market. Just marginally overbidding given an inactive policy of the opponent is enough to achieve an immediate reward in the market. Such price strategies by both agents imply non- cooperative behavior between firms.

In Proposition 3 we suggested that introducing a minimal amount of anticipation in the agents' pricing policy might invite agents to revise their myopic decision. Indeed the interaction scenarios that obtain outcomes from the basis of myopic optimal play do not replicate the process of learning by agents' decisions and hence are troubling to be prescriptive. Rather they at the best identifying the dynamic of a world where learning is irrelevant or the condition under which learning can or should stop. For example Maskin and Tirole (2001) study the Markov-perfect equilibria of games in which players are worried about predictions of future because the future does matter. Analogously in our model one party not only thinks "What happens to my payoff if I overbid?" but also "Will I be overbid by the opponent player after I have already overbid?" And "what is going to be my next reaction after my opponent's reaction?" The ability of real human agents to reflect on their behavior based on subsequent reactions to their decisions will distinguish them from "non-intelligent" myopic actors. In

order to test proposition 3, in the next section we investigate whether extending the foresight of agents will allow for some equilibrium price range to emerge.

2.3 The effect of contractual relationships on coordinative behavior

2.3.1 *Compensatory strategies*

In this section we allow the location of firms to be changed on the line market. The disputed area remains flexible as at the end of the previous section. In addition, processors face competition from one direction only. Assume each firm learns from its contracted suppliers when competitors try to negotiate supply contracts with them. They will try eventually to convince their contracted suppliers not to switch, for example, by means of supply contracts providing compensation. In this section we let there be some minimum margin Δ Agent *A* (*B*) might pay more to deter agent *B* (*A*) from seeking to acquire supply contracts of *A* (*B*). If firm *A* and *B* have set their price equal to *u* in *t=3*, firm *B* will be informed by its contracted suppliers that the opponent *A* is seeking to acquire supply contracts of *B*. Firm *B* will then increase its price up to the constant term $\Delta_{\boldsymbol{B}}$ (analogously one can presume the $\Delta_{\boldsymbol{A}}$ term as compensation parameter of agent *A*). The function $\Pi_A^{u_A}(B,[u_B,\Delta_{\boldsymbol{B}}])$ describes the utility of firm *A* if it decides to increase its price to u_A given the strategy profile of firm *B* consisting of price u_B and the compensation parameter $\Delta_{\boldsymbol{B}}$. Assume the strategy profile $\{u_B,\Delta_{\boldsymbol{B}}\}$ for agent *B* is given. Subject to the distance between firms *R* ($0<R<\infty$) and transport cost *t*, the possible responses of agent *A* to *B*'s compensation activity might be distinguished by the following 3 distinct cases: Either *A* will undertake a 'Competitive Pricing Strategy'($\prod_A^{u_B+\Delta_B}$).

The 'Competitive Pricing Strategy' (CP) of *A* grants a utility equal to[17]

[17] The upper boundary of the integral must guaranty that the milk delivery area does not exceed the market boundary *R*.

$$\Pi_A^{u_B+\Delta_B}(B,[u_B,\Delta_B]) = \left(\int_0^{Min\left\{R,\frac{P_A-u_B-\Delta_B}{t}\right\}} \left(P_A - u_B - \Delta_B - tr\right) dr \right) \left(u_B + \Delta_B\right)^{\varepsilon} \tag{2.2}$$

Or by knowing the compensation possibilities of agent *B*, *A* would abstain from acquiring new contracts and remain with the price u_A; In this case the 'Price-Matching Strategy' (PM) of *A* ($\Pi_A^{u_B}$) grants a utility equal to[18]

$$\Pi_A{}^{u_B}(B,[u_B,\Delta_B]) = \left\{ \begin{array}{l} \left\{ \int_0^{Min\left(R-Min\left\{R,\frac{P_B-u_B}{t}\right\},Min\left(R,\frac{P_A-u_B}{t}\right)\right)} \left(P_A - u_B - tr\right) dr \right\} + \\ 0.5 \left\{ \int_{R-Min\left\{R,\frac{P_B-u_B}{t}\right\}}^{Max\left\{R-Min\left\{R,\frac{P_B-u_B}{t}\right\},Min\left(R,\frac{P_A-u_B}{t}\right)\right\}} \left(P_A - u_B - tr\right) dr \right\} \end{array} \right\} u_B{}^{\varepsilon} \tag{2.3}$$

As a last alternative, *A* may decide to underbid agent *B* and act as the residual claimant of the market area (where agent *B* is non-active) by setting *A*'s monopsony price in this market. We name the latter strategy the 'Non-Competitive Pricing Strategy' (NCP) of *A* ($\Pi_A^{u_A^M}$).We assume that if the residual market area $R - (P_B - u_B)/t$ is smaller than the calculated optimal monopsony area of *A* in the previous section, $2P_A/((2+\varepsilon)t)$, then *A* will take the residual market as given and maximizes its utility within this area by its 'Non-Competitive Pricing Strategy' of *A*, hence if $R - (P_B - u_B)/t < 2P_A/((2+\varepsilon)t)$ we have:

$$\Pi_A^{u_A^M}(B,[u_B,\Delta_B]) = \left\{ \int_0^{R-Min\left\{R,\frac{P_B-u_B}{t}\right\}} \left(P_A - u_A^M - tr\right) dr \right\} \left(u_A^M\right)^{\varepsilon} \tag{2.4}$$

[18] The upper bound of the first integral not only foresees the market boundary *R* but also guaranties that in the case of no market overlap, the exclusive market area of *A* will cease at the point $Min\left(R,\frac{P_A-u}{t}\right)$ and in the case of a market overlap will cease at the point $R - Min\left(R,\frac{P_B-u}{t}\right)$. In the case of no market overlap, the upper and the lower bound of the second integral comply with each other.

It can be shown, that the monopsony price of agent A given the residual market area $R-(P_B - u_B)/t$ is

$$u_A^M = \frac{\varepsilon}{2(1+\varepsilon)}\left\{2P_A - t\left(R - \frac{P_B - u_B}{t}\right)\right\} \tag{2.5}$$

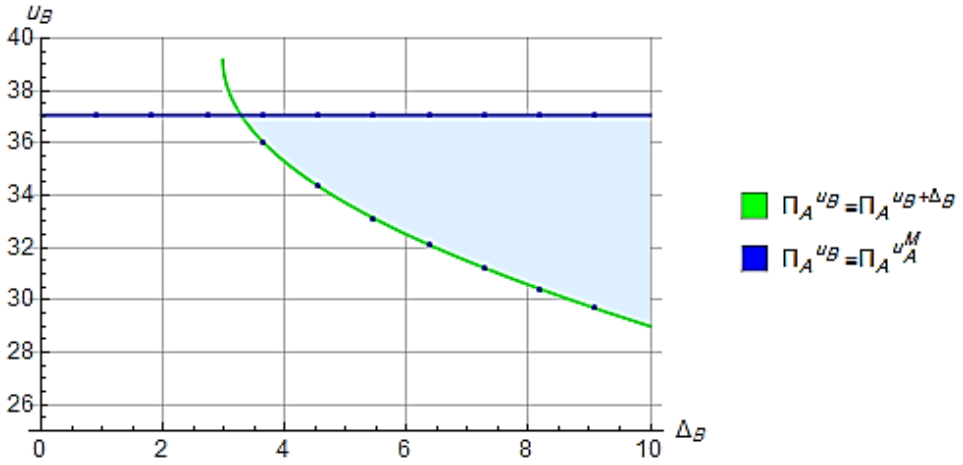
In contrast we assume that if the residual market area $R-(P_B - u_B)/t$ is larger than the calculated optimal monopsony area of *A*, then *A* will set its optimal monopsony price at its optimal Monopsony area as derived in section 2.2, hence if $R-(P_B - u_B)/t \geq (2P_A)/(2+\varepsilon)t$ the 'Non-Competitive Pricing Strategy' of *A* grants the following utility for A:

$$\Pi_A^{u_A^M}(B,[u_B,\Delta_B]) = \left\{\int_0^{\frac{2P_A}{(2+\varepsilon)t}}\left\{P_A - \frac{\varepsilon P_A}{(2+\varepsilon)} - tr\right\}dr\right\}\left(\frac{\varepsilon P_A}{(2+\varepsilon)}\right)^{\varepsilon} \tag{2.6}$$

Analogously one can derive the aforementioned equations subject to the distance *R* and transport cost *t*, for the best response of agent *B*. Setting each of the utilities $\Pi_A^{u_A^M}$, $\Pi_A^{u_B}$ and $\Pi_A^{u_B+\Delta_B}$ equal to each other and drawing the curves in a coordinate system will lead to phase diagrams similar to Figure 2.5. All points on the green curve depict the price and compensation combinations of agent *B* which render agent *A* to be indifferent between implementing PM and CP. On the other hand all points on the blue line depict the price and compensation combinations of agent *B* which render agent *A* to be indifferent between implementing a PM and NCP. Given that both firms have set their prices to *u* the blue area shows the region where PM is most beneficial for *A* knowing that he must at least face compensations by *B* equal to Δ. As we have seen, increasing the price is not improving one agent's overall payoff for certain ranges of compensation constraints by its competitor.

Figure 2.5: Optimal strategies of agent *A* based on *B*'s strategy profile *(u,Δ)*.

Note: The area in the upper side of the curve $\prod_A^{u_B} = \prod_A^{u_B+\Delta_B}$ depicts the dominance area of PM in contrast to CP. The area underneath the line $\prod_A^{u_A^M} = \prod_A^{u_B}$ depicts the dominance area of PM in contrast to NCP. The light blue shaded area depicts the dominance area of PM against other strategies. Agent *A* and *B*'s downstream market price = 75 and Transport cost and price elasticity of supply are set to 1. Distance between firms=50.



2.3.2 *Foresight based Nash prerequisites and conditions*

Agents in the reality might not decide in a simultaneous static game. The sequential interaction of agents in the milk market - featuring contractual relationships - causes each player to model its opponent by building beliefs about dynamic price-policies of its opponent. We propose that in such dynamic, interactive environments some information play a crucial role for an agent especially knowing about the *outside line* of its opponent. Let's define agent *A*'s outside utility line u_A^O as the price set by the opponent *B* that renders any infinitesimal overbidding pricing strategy δ not useful anymore for agent *A* and will invite *A* to survive in the residual market area by means of its NCP (recall proposition 1). At this point, the utility generated in the market area of agent *B*

constitutes per definition agent *B*'s *Outside Option*.[19] We define the outside utility line of agent *A* formally as follows (one may derive outside utility line of agent *B* and determine Outside Option of agent *A* in the same way):

$$\left| \begin{array}{l} \vartheta \leftarrow \left[\Pi_A{}^{u^* + \delta}(B, [u_B, \Delta_B]) < \Pi_A{}^{u_A^M}\left(B, [u_B, \Delta_B]\right) \wedge \delta \to 0 \right] \\ u_A^O = u^* if \begin{cases} if\ (u_B \geq u^*) : \vartheta = True \\ Else : \vartheta = False \end{cases} \end{array} \right| \tag{2.7}$$

Outside utility line of agents help them to model each-others' responses to its own decisions. According to the above mentioned definition, all points promising prices under the outside utility line of *A* will be perceived by opponent agent *B* as credible threat. We can define all points under the equation $u_A + \Delta_A \leq u_A^O$ analogously as *Leaving-Line* of agent *A* (*AL*). Respectively one may define u_B^O and *BL* as agent *B*'s outside line respectively *B*'s Leaving-Line. Through incorporating aforementioned two constraints for policy setting of agents we would like to investigate necessary and sufficient conditions that will establish a strategy profile consisting of prices and compensation-foresight *{(A,*$[u^{NE}, \Delta_A^*]$*)* & *(B,*$[u^{NE}, \Delta_B^*]$*)}* as Nash equilibrium of the spatial game in *t=3*. Let assume that both agents have set their price at the target price level u^{NE}. We first assume that by the prices u^{NE} both agents are better off than their non-competitive strategy profile. In other word *assumption 1* expressed in (2.8) is simply indicating that the equilibrium price is of interactive nature:

$$\exists u^{NE} \leftarrow \forall i, j \in \{A, B\}, i \neq j : \Pi_i^{u_i^M}\left(j, \left[u^{NE}, \Delta_j^*\right]\right) < \Pi_i^{u^{NE}}\left(j, \left[u^{NE}, \Delta_j^*\right]\right) \tag{2.8}$$

Assumption 1 might facilitate that neither of agents is convinced to be outside of market knowing the price set by the other. Given this, both agents either are seeking to undertake a CP or a PM strategy. For example, in Figure 2.7 all prices underneath the line "Assumption 1" fulfill the above mentioned properties for

[19] Note that if the outside utility line of *A* is less than Agent *B*'s monopsony price (or vice versa), then no competition emerges in the market.

both agents. Proceeding from assumption 1 the following conditions 1-4 provide necessary prerequisites for both rational parties not to have right incentives for establishing unilateral market by undertaking competitive policies.

Condition 1 and 2 are defined in 2.9 for *B* repectively *A*: (2.9)

$$\forall i, j \in \{A, B\}, i \neq j:$$
$$\exists \Delta u_i^* \rightarrow \left[\forall u_i \in \Delta u_i^* \exists \Delta_i^* \wedge \left[\begin{array}{l} u_i + \Delta_i^* \leq u_i^O \\ \wedge \Pi_j^{u_i + \Delta_i^*} \left(i, \left[u_i, \Delta_i^* \right] \right) < \Pi_j^{u_i} \left(i, \left[u_i, \Delta_i^* \right] \right) \end{array} \right] \right]$$

Condition 1 describes if *i*=*A*, then based on *B*'s deliberation how *B* might make its optimal pricing policy based on he models agent *A*'s responses to its own decisions. Deliberating to set any policy from price u_A onward, *B* must estimate the efficacy of setting a competitive price by considering the compensations he might face described by $\{u_A, \Delta_A\}$. The curve corresponded to equation $\Pi_B^{u_A + \Delta_A^*} \left(A, \left[u_A, \Delta_A^* \right] \right) < \Pi_B^{u_A} \left(A, \left[u_A, \Delta_A^* \right] \right)$, which we name Competing-Line of *B* (BC), can be drawn as a mathematical equivalent of *B*'s deliberation. *B* knows that *A* – given *B* setting a price higher than u_A and lower than $u_A + \Delta_A^*$ will offer compensation to its contracted partners. Indeed based on *A*'s outside utility line, *A* can reasons that price compensations render a utility higher than the utility of being just a residual claimant of the outside market area. As long as *B*'s price does not surpass *A*'s outside utility line, compensation by *A* is perceived by *B* as a credible threat. In addition if condition 1 holds, then *B* could achieve a higher utility equal to $\Pi_B^{u_A} \left(A, \left[u_A, \Delta_A^* \right] \right)$ by PM at u_A, compared to a CP price higher than $u_A + \Delta_A^*$. Briefly explained, condition 1 allows finding the price range in which agent *B* in convinced of setting the price to u_A rather than undertaking any feasible CP strategy. Condition 2 might be understood analogous to condition 1 and describes agent *A*'s decision model based on *A*'s model of *B*'s behavior. Note that establishing mutual coordination fails if (at least) one party feels - in contrast to the implied knowledge in above mentioned conditions – that he can attain larger margins by setting its price e.g. equal to the outside utility line of his

opponent. Such a party might not commit to mutual coordination and withhold cooperation. Condition 3 verifies whether the price spectrum of aforementioned interest of firms in condition 1 and 2 overlap.

Condition 3:

$$\Delta u_A^* \cap \Delta u_B^* \neq \varnothing \tag{2.10}$$

We name the price span $\Delta u_A^* \cap \Delta u_B^*$ the *permissible price range*, since it provides the incentives for both firms to be convinced regarding a coordinative pricing policy instead of pursuing unilateral competitive policies. Lastly, one must take into consideration that assembling bilateral accommodation will rely on agents' beliefs regarding the opponent's simultaneous behavior in the equilibrium. Condition 4 prescribes the prerequisite knowledge for establishing coordination in permissible price range:

Condition 4:

4.1. A knows that B knows A's behavior in line with condition 1.

4.2. B knows that A knows B's behavior in line with condition 2.

4.3. Both agents act consistent with the beliefs (4.1) and (4.2).

Condition 4 implies that knowing more about the decision model of each agent does not alter the decision of either of agents regarding efficient coordination implied in condition 3. As it is shown price-matching equilibria can potentially emerge based on agents' subjective belief regarding the competitor agent's Leaving-Line. Conditions 1-4 might serve to show how sequential interaction nature of the game from proposed u^{NE} onward can distort one-step look ahead myopic policies of processor agents and alter the nature of decision making by pricing towards what my overbidding does and what might come next by the rival agent after I overbid. In such a circumstance outside utility lines of firms u_B^O and u_A^O can potentially serve as reference points reminiscent of dynamic practice of opponent regarding its compensations by adapting its prices after he knows about decisions of rival. With regard to Non-existence of Nash equilibrium in the game each firm will necessarily try to give a best myopic response by interfering in market e.g. through negotiations with the suppliers after he learns that its

opponent has achieved new arrangements with suppliers in the market. Conditions 1-4 persuade each party that the utility generated in the market area of the agent by setting its price to outside utility line of the rival (and allowing to reach its Outside Option) will not grant a higher payoff compared to a coordinative policy together with opponent player in permissible price range. Therefore whereas these conditions link the relative usefulness of competitive strategy of agents to the outside line of the rival, we must ensure that the Outside Option of agents is the maximum utility they can achieve by undertaking any unilateral pricing policy.[20] If such assumption holds, estimation of price-matching values in conditions 1 and 2 with reference to outside utility lines is sufficient for making agents aware of the maximum achievable utilities without coordination (based on the behavioral model of the opponent). Therefore *assumption 2* is introduced to consolidate conditions 1 and 2. Let's name the utility one agent by setting its price to a price *u* –ex post of opponent's next best myopic response to u- achieves the agents Withhold-Cooperation payoff. One might understand the Withhold-Cooperation pay-off of one firm in our dynamic game similarly with regard to the notion of Max-Min payoff (Neumann, 1928). However note that by insisting on its Withhold-Cooperation policy, each agent presumes rational behavior model of its opponent. Hence we presume that one agent selects its Withhold-Cooperation policy unilaterally in a first stage and by doing so it taught the opponent to adapt rationally in the following stage. Assumption 2 shall certain that withhold-cooperation payoff of both agents coincide with their Outside Option:[21]

$$u_i^O = \underset{u}{ArgMax}\,\Pi_i^{Withhold-Coop}(u_i), i, j \in \{A, B\} \tag{2.11}$$

We formulate thereby agent *A*'s Withhold-Cooperation payoff according to the utility *A* by setting its price to *u* –ex post of *B*'s next best myopic response to *u*-

[20] For example agents in conditions 1 and 2 just manage to sight outside utility line of the opponent and don't take into considerations whether payoff of a firm can exceed its Outside Option if the party were underbid by competitor.

[21] As it is aforementioned Outside Option of one agent corresponds to the point where he has set its price equal to Outside utility line of the rival.

archives. One might define the agent *B*'s Withhold-Cooperation payoff in the same way:

$$\Pi_A^{Withhold-Coop}(u_A)=\begin{cases}0 \Leftarrow u_A < u_B^{opt} \wedge R \le R_B^{opt} \\ \int_0^{Min\left(\frac{P_A-u_A}{t},R-\frac{2P_B}{(2+s)t}\right)} (P_A-u_A-tr)dr \Leftarrow u_A < u_B^{opt} \wedge R > R_B^{opt} \\ \int_0^{Min\left(R-Min\left(R,\frac{P_B-u_A}{t}\right),Min\left(R,\frac{P_A-u_A}{t}\right)\right)} (P_A-u_A-tr)dr \Leftarrow u_B^{opt} \le u_A < u_B^{O} \\ \Pi_A^{u+\Delta_B}(B,[u_A,0]) \Leftarrow u_B^{O} \le u_A\end{cases}$$

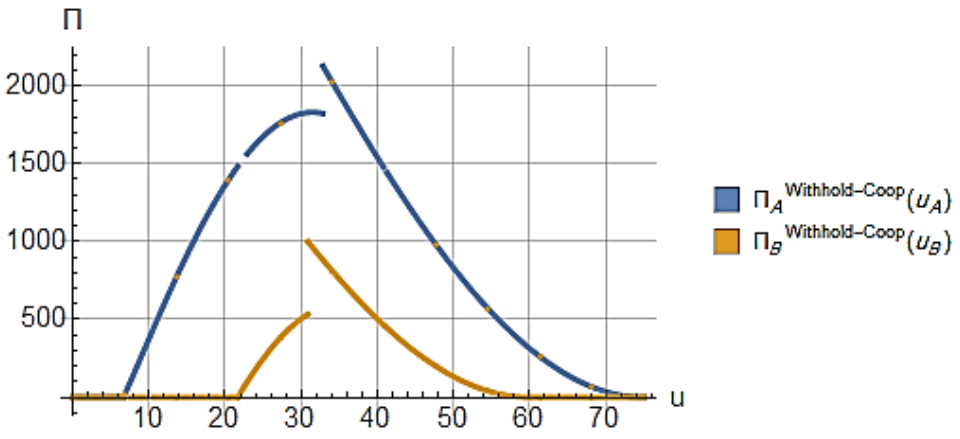
(2.12)

The upper term in equation 2.12 implies that if *A* sets a price lower than agent *B*'s optimal monopsony price (section 2.2.1) agent *B* will set just its optimal monopsony price. Hence, agent *A*'s Withhold-Cooperation payoff turns to be 0 if the market area *R* is not larger than *B*'s optimal monopsony area. The second term means if the market area *R* is larger than *B*'s optimal monopsony area then *A* will earn some payoff in the residual market area by setting its price to *u*. The third term in equation 2.12 describes circumstances where *A* might have set its price equal to *u*, afterwards *B* will slightly overbid *A*. The lowest term indicates the *A*'s payoff if *A* is undertaking pricing policies above *B*'s outside utility line. In this case *A* will just gain a utility which amounts to its competitive pricing policies including the compensation of *B* (Δ_B) amounting to zero. Figure 2.6 shows the fulfilment of assumption 2 in an exemplary market setting due to Withhold-Cooperation payoff of agents. Discontinuous nature of the function can reveal this fact clearly. Please note that assumption 2 remains in all experimentation instances in our paper true. To sum up when assumption 1-2 hold true, both agents get deprived of any incentives for targeting relatively unilateral through conditions 1-4 margins instead of mutual beneficial coordination. A Nash equilibrium emerges comprising prices and compensation-foresight consisting of

the strategy profile (u^{NE}, Δ_A^*) for *A* and (u^{NE}, Δ_B^*) for *B*. In equilibrium no party believes it can achieve a higher competitive payoff rather than its price-matching utility if coordination gets revoked. Although Outside Options of firms remain out of equilibrium, the outside lines of firms remain a credible threat for the opponent.

Figure 2.6: Withhold-Cooperation payoff of agents based on opponent's behavior model.

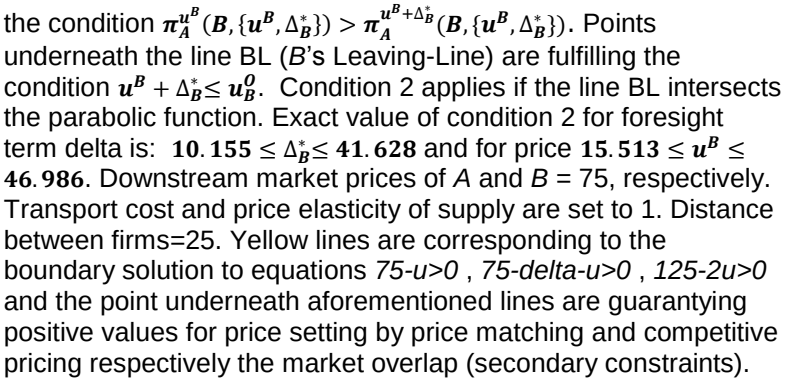
Note: Down-stream Market prices are assumed 75 & 60 for agent *A* and *B*, respectively. Transport cost set to 1. Price elasticity of supply = 0.25 and distance between firms = 53.33. $\boldsymbol{u_B^O}$ =32.8336 and $\boldsymbol{u_A^O}$ =30.8801 represent maximum points in both panels.



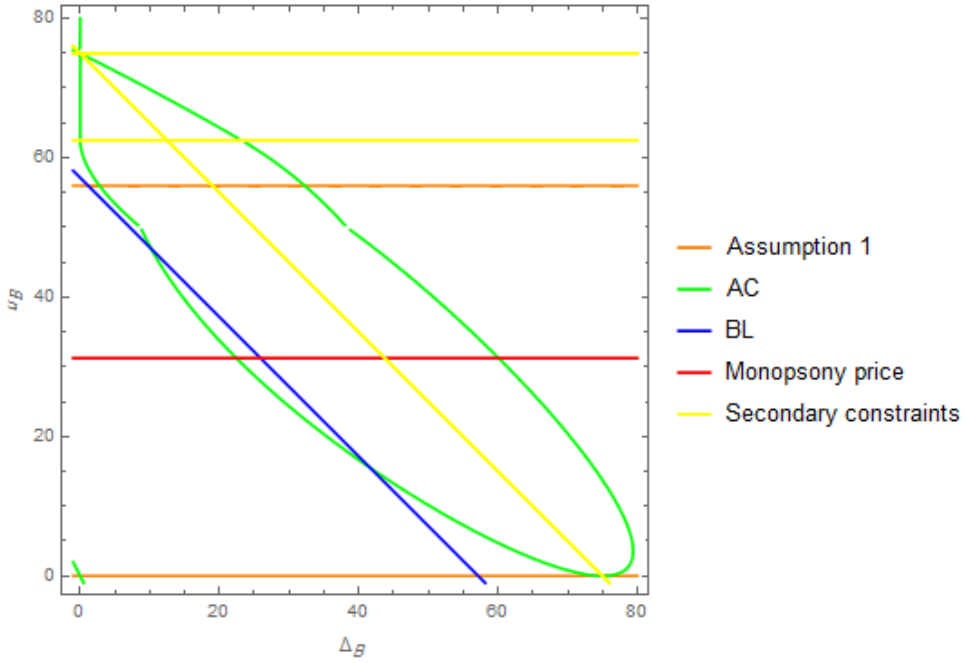
2.3.3 *Equilibrium in an exemplary market setting*

So far we have set necessary and sufficient conditions for establishing a permissible price range. With the following illustration we show how strategy profiles fulfilling conditions 1-4 may lead to the emergence of Nash equilibrium in our exemplary market setting. In Figure 2.7, points under the line BL (agent *B*'s Leaving-Line) represent the extent to which overbidding yields higher utilities for *B* compared to being just the residual claimant of the remaining market area. Points within the curve AC (*A*'s Competing-Line) depict the region where further overbidding for *A* ceases to be useful. In the area between the lines, condition 2 is true i.e. price u^B is the best response of agent A given strategy profile $\{u^B, \Delta_B^*\}$ of B.

Figure 2.7: Illustration of the credible threat by agent B and accommodation by A.



Note: Points within the area of parabolic function AC are fulfilling the condition $\boldsymbol{\pi_A^{u^B}(B,\{u^B,\Delta_B^*\}) > \pi_A^{u^B+\Delta_B^*}(B,\{u^B,\Delta_B^*\})}$. Points underneath the line BL (*B*'s Leaving-Line) are fulfilling the condition $\boldsymbol{u^B + \Delta_B^* \leq u_B^O}$. Condition 2 applies if the line BL intersects the parabolic function. Exact value of condition 2 for foresight term delta is: $\boldsymbol{10.155 \leq \Delta_B^* \leq 41.628}$ and for price $\boldsymbol{15.513 \leq u^B \leq 46.986}$. Downstream market prices of *A* and *B* = 75, respectively. Transport cost and price elasticity of supply are set to 1. Distance between firms=25. Yellow lines are corresponding to the boundary solution to equations *75-u>0* , *75-delta-u>0* , *125-2u>0* and the point underneath aforementioned lines are guarantying positive values for price setting by price matching and competitive pricing respectively the market overlap (secondary constraints).



By having equal prices and assuming that both agents come up with the same situation, agent *B* can also induce that there would emerge no beneficial market access at the end of the game without coordination with the other firm *A*. A price-matching strategy within the permissible price range is rational for both parties given the beliefs held by agents about opponent's dynamic policy.

2.4 Importance-of-space, inelastic supply and disparity between firms

Coordination may not be easily sustained. Shared market access may not be applied by agents in Nash equilibrium if, for example, either party lacks the incentive to cooperate (condition 1 and 2) or the areas of interest don't intersect (condition 3). The spatial shape of the market and differences between the compensation flexibilities of firms may cause a firm to not being able to foresee the difficulty faced if it starts an overbidding policy from the permissible price range onward. Another issue is inelastic supply in the short run (Gardner, 1992). In fact farmers cannot suddenly supply more milk if the price goes up. This limited flexibility creates a relatively inelastic supply of milk with respect to price in the short term.

In order to test the robustness of our results in section 2.3 we would like to experiment possibility of processors' cooperation in different market structures. The key variable indicator in predecessor studies of spatial competition is the so called *importance-of-space* as a measure for competitiveness of a market. Importance-of-space (*I*) can be calculated through multiplying transport costs (*t*) by distance between competitors (*R*) divided by net value of product being sold at the downstream market (ρ), i.e. $I = t * R/\rho$ (Alvarez et al., 2000). As I might increase, competition between the firms diminishes to the point where eventually they are spatially isolated monopsonies. Hence we presume I as the explanatory variable of our simulation study and investigate the nature of spatial competition by exogenously switching the location of firms toward each other.[22] The relation between firm locations and collusive behavior also has been studied well analytically in the literature of horizontal product differentiation. The results obtained in a smaller body of literature shows that a smaller firm dispersion is more likely to sustain tacit collusion. For example the critical discount factor required to sustain collusion in Gupta and Venkatu (2002) and Matsumura and Matsushima (2005) monotonically decreases as firms are located close together. However the majority of studies (Chang, 1991; Häckner, 1995 and Miklós-Thal, 2008) suggest that the relationship between product differentiation and collusive behavior is robust. The results of aforementioned papers are relying on the perception that it is difficult to sustain collusion for strong substitutes (e.g. when the location of firms in space is interlocked). This is because the gains from defection become larger. In contrast, a higher degree of horizontal differentiation should help firms to sustain collusion because it renders deviations less profitable. Note that all above papers prescribe a reversion to an infinitely repeated static Nash equilibrium, right after cheating by any single party, whereas the obtained results in our model are relying on compensatory actions of agents in the dynamic game described in section 2.2.

In order to design experiments in this section let's first take the optimal monopsony area of firm A derived in section 2.2.1 ($S = 2P_A/((2+\varepsilon)t)$) as a

[22] For access to the code please see online appendix 1:
(http://www.ilr.uni-bonn.de/agpo/staff/khalili/khalili.zip)!

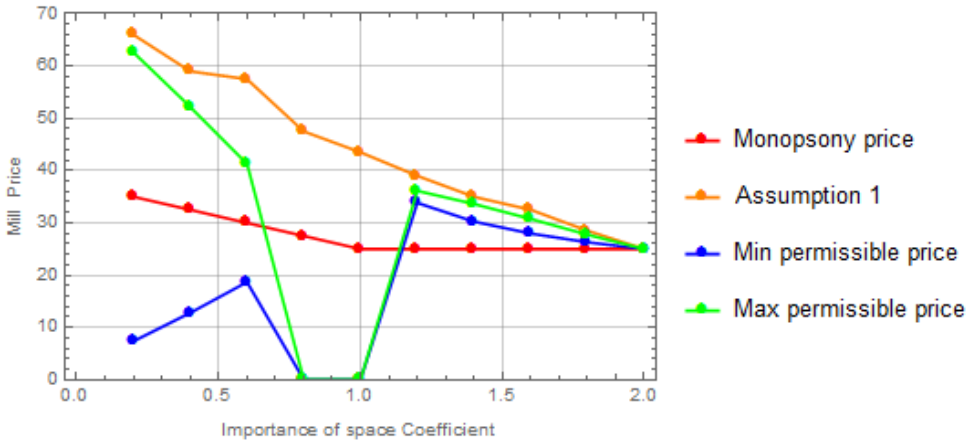
measure and simulate the market by setting the distance between firms R equal to $\varphi \times S$ where φ is a coefficient representing the importance-of-space I. In the second step we replicate the first simulation by setting the term price elasticity of supply to a low level equal to 0.2. In the third step we illustrate an example for firms having different prices of end-products at downstream market.

Figure 2.8 depicts the permissible price range in an exemplary market setting by incrementing the parameter φ form the situation where firms locations are almost interlocked (φ=0.2) up to the point $2 \times S$ where processors' location imply 2 monopsony areas. Results show that - with the exception of 2 intermediary market structures (φ=0.8, φ=1) – equilibrium conditions 1-4 apply in the majority of market conditions.[23] Players mutually learn not to give a best myopic response in the less and high competitive market structures. The lower the importance-of-space, agents are supposed to learn better from each other by building price conjectures about the opponent and consequently more likely settle the price dispute by mutual concession. This can be confirmed by looking at the difference between the maximum and minimum thresholds of the permissible price range in Figure 2.8. The greater the spatial distance the more restricted is the range of coordination. By gradually increasing the importance-of-space, the utility of NCP for firms become more attractive. However, establishing cooperative behaviors in markets characterized by intermediate structures seems to be awkward. This outcome confirms the analytical contribution in the literature on horizontal product differentiation that a larger spatial dispersion of firms might support collusive behavior, since a higher product differentiation means a greater Outside Option for competing firms in our model, but in the same time we show that by adding the effect of institutional factors -as a major caveat to understanding real market interactions- sometimes smaller spatial dispersion of firms might also encourage tacit collusion. The reason is that the degree of foresight-based reasoning by learning agents is higher the lower the importance-of-space.

[23] Note that this means in the market structures indicating φ=0.8 and φ=1 PM coordination fails to be established as parties have greater withhold cooperation payoffs!

Figure 2.8: Interaction of firms by changing the explanatory variable importance-of-space with a unitary price elasticity of supply.

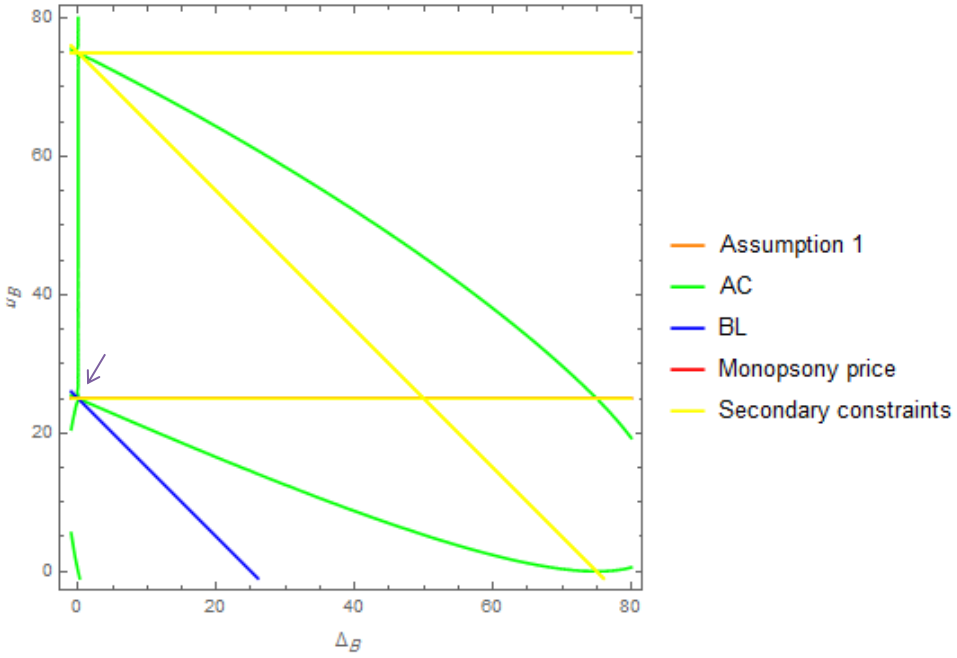
Note: The parameter $\varepsilon = 1$. Downstream market prices of $A = 75$ and $B = 75$ respectively Transport cost = 1.



By increasing φ, the permissible price range shrinks and is finally confined to the only point promising the equilibrium price range at $\varphi = 2$, i.e. to the optimal monopsony price derived in section 2.2.1. Figure 2.9 shows how by setting the coefficient of importance-of-space to its extreme value $\varphi = 2$ the interests of firms regarding price coordination and setting the Monopsony price coincide in the non-competitive market structure.

Figure 2.9: Intersection of outside utility of firms with monopsony price line in a market with highest importance-of-space.

Note: The parameter $I = t^*R/\rho = 4/3$. Assumption 1 matches the Monopsony price line and the secondary constraints regarding market coincide. Exact value for condition 2 is: $\mathbf{0} \leq \Delta_B^* \leq \mathbf{0}$ and $\mathbf{25} \leq \boldsymbol{u^B} \leq \mathbf{25}$.



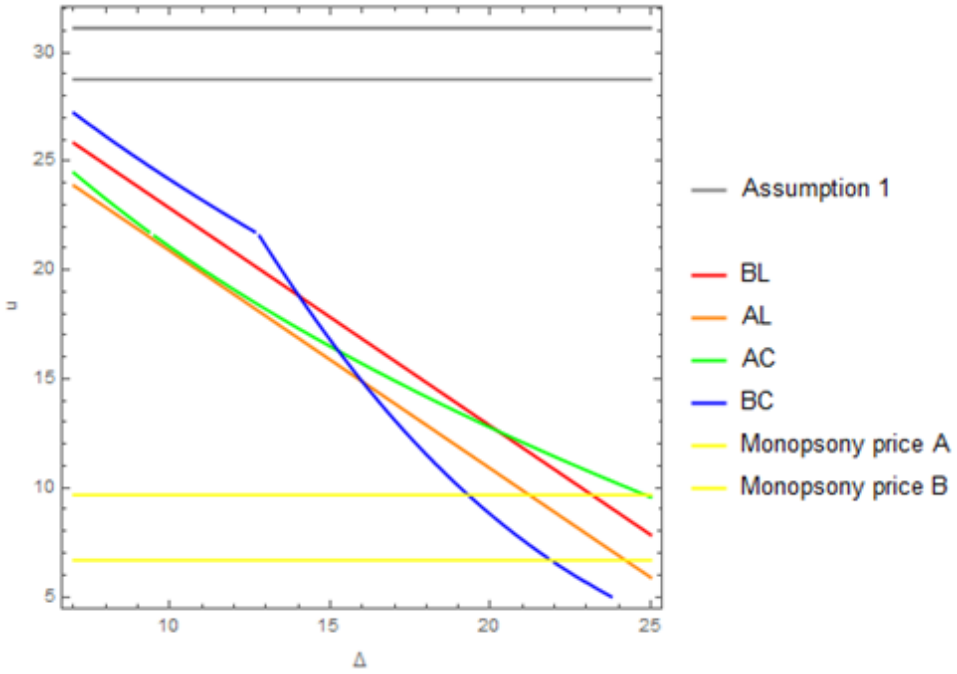
We repeated the experiment of changing importance-of-space subsequently with a low price elasticity of supply (ε= 0.2). The obtained results are very similar to those observed in Figure 2.8. With only one exceptional intermediate market structure (φ=1), equilibrium conditions 1-4 apply in a broad range of market structures. However, the lower price elasticity of supply dampened the incentive of firms for setting higher prices and consequently pushed down the lower threshold of permissible price range towards zero. Note that the minimum threshold would converge to the sunk cost of supplier farms in reality if we would have included this factor in the model.

We investigate finally how rational agents can learn to trust their co-player and pursue the common goal of achieving better utility even though the firms have asymmetric attributes such as different processing costs or different selling prices at downstream market. Figure 2.10 shows an example for firms' decisions involved in such situation. Downstream market prices are assumed 75 & 60 for agents A and agent B, respectively. The permissible price range without incentives for withholding cooperation is obtained within the price range 12.06 $\leq u^{NE} \leq$14.88. Note that whether the equilibrium knowledge (or beliefs) of agents emerge based on sharing knowledge through communication or tacit collusions

without communication more strongly support this equilibrium is not incorporated in our model.

Figure 2.10: Illustration of Nash equilibrium by asymmetric firms.

Note: Exact value for condition 1 in the intersection area of BC (*B*'s Competing-Line) and AL (*A*'s Leaving-Line): $\mathbf{15.99} \leq \Delta_A^* \leq \mathbf{28.81}$ and $\mathbf{2.06} \leq \boldsymbol{u^A} \leq \mathbf{14.88}$. Exact values regarding foresight and price of condition 2 is in the intersection area of AC (A's Competing-Line) and BL (B's Leaving-Line): $\mathbf{3.76} \leq \Delta_B^* \leq \mathbf{20.22}$ and $\mathbf{12.60} \leq \boldsymbol{u^B} \leq \mathbf{29.07}$. Equilibrium price range $\boldsymbol{u^{NE}}$ where condition 3 applies is limited within $12.06 \leq \boldsymbol{u^{NE}} \leq 14.88$. Down-stream Market prices are assumed 75 & 60 for agent *A* respectively *B*. Transport cost set to 1. Price elasticity of supply = 0.25 and distance between firms = 53.33.



2.5 **Conclusion**

Undesirable system characteristics such as endless price wars are reported by classic pricing models (Shubik, 1980) and are discussed in spatial competition (Schuler and Hobbs, 1982). The significance of our model is in revealing the role of Outside Options of firms in the emergence of such patterns but also how by

introducing a minimal amount of anticipation in the agents' pricing policy may cyclic price wars move towards cooperation.

We additionally propose that contractual relationships in the milk market expedite the described foresight-based decision making. In contrast to spot markets where market participants more likely act like myopic responders, firms in long-term contractual relationships might try to re-negotiate with their suppliers after they become aware of opponents' efforts to acquire their supply aiming to keep them from switching.

One additional crucial implication of our model is the relationship between firms' location and collusive behavior. This problem is studied analytically in the literature of horizontal product differentiation. The majority of studies suggest that the relationship between higher degrees of product differentiation and collusive behavior is robust. Our model approves that a larger spatial dispersion of firms indicates less competition in the market as it grants a larger Outside Option for firms. However we add the effect of *institutional factors* as a major caveat to understanding real market interactions. We propose that advantages of strategic thinking for players within the system are more, less the importance-of-space is and hence a smaller dispersion of firms might support tacit collusions too.

2.6 References

Alvarez, A.M., Fidalgo, E.G., Sexton, R.J. and Zhang, M. (2000). Oligopsony Power with Uniform Spatial Pricing: Theory and Application to Milk Processing in Spain. European Review of Agricultural Economics, Vol. 27(3), pp. 347-364.

Beckmann, M.J. (1976). Spatial Price Policies Revisited. Bell Journal of Economics 7, pp. 619-630.

Boysen, O. and Schröder C. (2006). Economies of Scale in der Production versus diseconomies: on the structural change within the dairy sector. Agrarwirtschaft 55 Heft 3, pp. 152-167.

Bundeskartellamt. (2009). Inquiry Milk sector. Bonn, Germany: Federal cartel office.

Bundeskartellamt. (2012). Inquiry Milk sector. Bonn, Germany: Federal cartel office.

Capozza, D.R. and Van Order, R. (1978). A Generalized Model of Spatial Competition. American Economic Review, Vol. 68(5), pp. 896-908.

Chang, M.H. (1991). The effects of product differentiation on collusive pricing. International Journal of Industrial Organization 9 (3), pp. 453–469.

Dasgupta, P. and Maskin, E. (1986). The Existence of Equilibrium in Discontinuous Games. Applications. Review of Economic Studies 53, pp. 27-41.

Durham, C.A., Sexton R.J. and Song J.H. (1996). Spatial Competition, Uniform Pricing and Transportation Efficiency in the California Processing Tomato Industry. American Journal of Agricultural Economics 78, pp. 115-125.

European Commission - DG Agriculture and Rural Development. (2013). AGRI-2012-C4-04 - Analysis on future developments in the milk sector.

Fudenberg, D. and Levine, D. K., The theory of learning in games. (1998). Cambridge, MIT Press.

Gardner, B. (1992). Changning Economic Perspectives on the farm problem. Journal of Economic Literature 30, pp. 62-101.

Graubner, M., Balmann, A. and Sexton, R.J. (2011b). Spatial Price Discrimination in Agricultural Product Procurement Markets: A Computational Economics Approach. American Journal of Agricultural Economics, Vol. 93(4), pp. 949-967.

Graubner, M., Koller, I., Salhofer, K. and Balmann, A. (2011a). Cooperative versus Non-cooperative Spatial Competition for Milk. European Review of Agricultural Economics. European Review of Agricultural Economics, Vol. 38(1), pp. 99-118.

Greenhut, G. (1981). Spatial pricing in the USA, West Germany and Japan. Economica 48, pp. 79-86.

Greenhut, M.L. and Ohta, H. (1972). Monopoly Output under Alternative Spatial Pricing Techniques. American Economic Review 62, pp. 705-713.

Gronberg, T. and Meyer, J. (1981). Competitive Equilibria in Uniform Delivery Pricing. The American Economic Review, Vol. 71(4), pp. 758-763.

Gupta, B. and Venkatu, G. (2002). Tacit collusion in a spatial model with delivered pricing. Journal of Economics 76 (1), pp. 49-64.

Häckner, J. (1995). Endogenous product design in an infinitely repeated game. International Journal of Industrial Organization 13 (2), pp. 277–299.

Hart, O. (1995). Firms, Contracts and Financial Structure. Oxford: Oxford University press.

Hotelling, H. (1929). Stability in Competition. Economic Journal 39, pp. 41-57.

Huber, A. (2007 a). Entwicklungstendenzen in den Milcherfassungsgebieten deutscher Molkereiunternehmen. Technische Universität München.

Huber, A. (2009). Milcherfassungsgebietskarte deutscher Molkereiunternehmen. Technische Universität München.

Huck, P., Salhofer, K. and Tribl, C. (2006). Spatial Competition of Milk Processing Cooperatives in Northern Germany. International Association of Agricultural Economists Conference.

Lösch. (1954). The Economics of Location. New York: Yale University Press.

Maskin, E. and Tirole, J. (2001). Markov perfect equilibrium. Journal of Economic Theory, Vol.20, pp. 191-215.

Matsumura, T. and Matsushima, N. (2005). Cartel stability in a delivered pricing oligopoly. Journal of Economics 86 (3), pp. 259–292.

Miklós-Thal, J. (2008). Delivered pricing and the impact of spatial differentiation on cartel stability. International Journal of Industrial Organization 26, pp. 1365–1380.

Neumann, J.V. (1928). Zur theorie der Gesellschaftsspiele. Mathematische Annalen,100(1).

Sexton, R. J. (1990). Imperfect Competition in Agricultural Markets and the Role of Cooperatives. American Journal of Agricultural Economics, Vol.72 (3), pp. 709-720.

Sexton, R. J. (2012). American Journal of Agricultural Economics, Volume 95, Issue 2, 1 January 2013, pp. 209–219.

Schuler, R. E. and Hobbs, B. F. (1982). Spatial Price Duopoly under Uniform Delivered Pricing. Journal of Industrial Economics 31, pp. 175–187.

Tribl. C. (2012). Spatial competition of food processors in pure and mixed markets under uniform delivered pricing. Germany: Submitted doctoral thesis, Technical University of Munich.

United states accountability office. (2004). Dairy industry, Information on milk prices, factors affecting milk prices, and dairy policy options. GAO.

Van Damme, E. (1989). Stable Equilibria and Forward Induction. Journal of Economic Theory, Volume 48, Issue 2, 48, pp. 476-496.

Chapter 3
Rational and Convergent learning in Multi-agent spatial markets

Abstract. Many suppliers, costly transport of raw products and relatively few couple of processors often characterize agricultural procurement markets. Figuring out the pricing policy of learning agents in such markets has been a recent issue of the recent computational and agricultural economics' literature. Yet learning in spatial systems requires algorithms that are scalable to large number of agents and can be implemented with minimal knowledge about the actions of other agents. Most proposed Multi-agent learning algorithms in the literature fail one or both of these criteria. Our research in this paper is set out to develop an operational algorithm in order to lead rational agents to adapt their policy in large-scale and dynamic strategy spaces with modest computational effort. We introduce a new learning algorithm, Hierarchical Policy Hill Climbing (H-PHC) and examine our algorithm with respect to rationality and convergence, two desirable properties for a learning system from the literature of MASs.

Keywords: Agricultural spatial markets, Reinforcement Learning, ABMs, Oligopsony

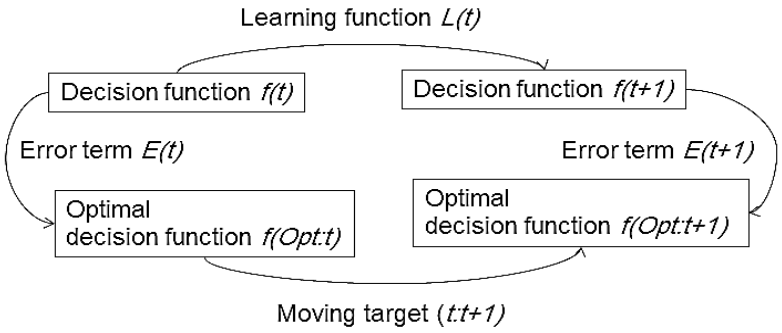
JEL classification codes: **C63, C70, D43, L13, L40**

3.1 Introduction

Price formation in agricultural procurement markets is a complex dynamic process with multiple agent interaction. Consider the case of a dairy processor seeking to maximize its profit in a raw milk procurement market. On the raw milk input side, farmers and competitors are distributed in space around processors'

locations. Processors compete with each other for milk delivery contracts with farmers. Each prospective farmer may get a price bid from each potential buyer and selects its business partner and production level based on offered bids. From a classical economic point of view, one may predict that in such systems where the goods, here the raw milk are homogeneous, the offered bids will continuously shoot up, until the milk delivery prices equal the value of products being sold at downstream market net of processing costs. However, consider the spatial feature of raw milk markets, where varying the distance to each processor changes the buyers willingness to pay or what farmers ask.

Figure 3.1: Learning problem in MASs is one of a moving target, adopted from Vidal (1998).



In the described context, each processor agent needs to dynamically keep up with the changes in the behavior of other agents. Finding the optimal pricing policy of processor agents among multiple, strategically interacting agents is not a trivial task. This problem is addressed as one of a "moving target" problem in the literature of MASs (Vidal, 1998). As depicted in Figure 3.1, a learning agent might try to learn its optimal decision function at time *t* subject to some cognitive error term *E(t)*. In interactive systems, each agent provides an effectively non-stationary environment for the other agents. In such situation it is far from certain that an agent keeping up with the changes in the environment up to time *t* has an expected error diminishing or converging to zero at time *t+1*. Because the optimal decision function in *t+1* may have moved to a new – in time *t* unforeseeable – point. In fact, agents are continuously exposed to new, incomplete information (and or knowledge) about the preferences and rationality of the other agents in the

game. This turns the task of computing a best response function complex (Parkes et al., 1997).[24] The complexity of many tasks arising in such interactive settings makes them difficult to solve with simplified agent behaviors. In this context, agents must instead discover a solution relying on their own knowledge, by learning.

The source of the instability problem spoiling the search procedure of agents in Multi-agent, interactive settings might be twofold. On one hand the actions of one agent strongly and frequently affect the plans of other agents. Hence when performance of an agent improves (gets worse), it is not necessarily clear whether the improvement (deterioration) is due to an altering in that agent's own behavior or a negative (positive) change in the opponent's behavior (Rosin and Belew, 1995).

On the other hand, pathological cyclic dynamics could emerge inherently from attributes of payoff matrices (Luke and Wiegand, 2002). For example, a major problem encountered in the current literature on spatial competition is the nonexistence of pure-strategy Nash equilibria in competitive models due to discontinuous best response functions (Dasgupta and Maskin, 1986; Beckmann, 1973; Schuler and Hobbs, 1982). In such cases, cyclic price wars take the place of Nash equilibrium. For example, if agents *A* and *B* would play a 'Rock-Scissors-Paper' game and agent *A* picks Rock, then agent *B* will pick Paper as a best myopic response. Agent *A*, then adapts to Scissors, causing agent *B* to switch to Rock. Then agent *A* adapts to Paper and agent *B* will consequently switch to Scissors. Then agent *A* adapts again to Rock. The same cycle will be again be reinitiated.

A significant part of the research on Multi-agent learning develops learning techniques trying to deal with the moving target problem (Busoniu et al., 2010). A major challenge in the presence of multiple players is however the curse of dimensionality caused by the exponential growth of the discrete state-action space

[24] Indeed standard game theory assumes that the rationality and preferences of all the agents is common knowledge (see Binmore, 1987).

(Busoniu et al., 2010).[25] The majority of algorithms are particularly suited to deal with small games (Tuyls et al., 2006) and only few single-agent learning tools offer the appropriate capability in the context of multiple learning agents.[26]

It is in ABM designers' best interest to consider methods, which obviate the need for agents to acquire excessive information e.g. deep knowledge about other agents and still coordinate well with environment (Durfee, 1995).

In this paper, we introduce an adaptive dynamic pricing model of learning in large-scale strategy spaces, which comprises hierarchically minded learning agents. The hierarchical learning approach allows players to pursue their learning goal by limiting the strategic scope of their learning subject to an evolving environment. Rational agents in reality guide their decisions by hierarchical elaboration rather than undertaking exhaustive enumerative searches over the whole decision space. This avoids that agents change their policies in a completely arbitrary fashion and reduces the dimensionality of interaction.

The remainder of the paper is organized as follows: After reviewing the relevant literature in section 3.2, section 3.3 provides offer a more precise insight to the H-

[25] Just assume four dairy processors i = 1, …4) offering bids' range u_i ($0<u_i<100$) and being positioned in the center of the four quadrants of some landscape. The total size of interaction space (by discretizing of u_i to predetermined increments having discrete values of 0.05) amounts 1.6e+13.

[26] A single learning approach is e.g. the no-regret algorithm (Cesa-Bianchi and Lugosi, 2006) where agents choose actions to minimize the regret of their choices based on their payoff history in previous rounds of the game. Even though the no-regret algorithm does not require information about other agents' actions, discretization of decision space by numerous agents causes this algorithm to be too slow to converge in large systems (Kash et al., 2011). Moreover, the convergence is not guaranteed in general settings. Some alternative methods are offered able to deal with restricted classes of games. For example, Kash et al. (2011) propose the Stage-algorithm. This algorithm delivers improved performances –compared to aforementioned methods- only for specific environments with countable many agents. Lack of scalability, i.e. functionality when increasing the size of discretization, characterizes other multi-agent learning algorithms, e.g. fictitious play (Fudenberg and Levine, 1998) or the experience-weighted attraction model (Camerer and Ho, 1999).

PHC mechanism. In section 3.4, we test our algorithm in a variety of spatial games. In section 3.4 we examine the H-PHC interaction model under markets with no Nash equilibrium as well as markets of non-spatial games. The final section includes some suggestion to improve the algorithm and final conclusions

3.2 **Literature context**

Basically achieving the learning goal of agents playing in MASs is strongly dependent on the agent's *degree of awareness* (Busoniu et al., 2010). Typically, other agents are implicitly or explicitly recognized and 'modelled' as entities having their own objectives and intentions. Based on their degree of awareness, either agents will learn just to correlate actions with rewards regardless of other agents' actions, or they will try to learn to predict the expected actions of others and use these predictions along with knowledge of the problem domain to determine their actions (Vidal, 2010). An example for an opponent *independent* algorithm is the Minimax-Q algorithm (Littman, 1994). The advantage of such algorithm is that it fulfils the safety criteria proposed in (Powers and Shoham, 2005).[27] The drawback of Minimax-Q is if the opponent is playing a suboptimal strategy, the algorithm doesn't learn to adapt. An example for an opponent *aware* algorithm, which takes the action of its opponent explicitly into consideration, is Wolf-PHC algorithm (Bowling and Veloso, 2001). For instance, Bowling and

[27] Diverse evaluation criteria are proposed in the literature to assess performance of learning agents in MASs. Criteria mentioned in Powers and Shoham (2005) are 'targeted optimality', 'compatibility' and 'safety'. 'Targeted optimality' implies that agent applying the algorithm achieves a certain minimal threshold of average payoff against a member of the selected set of opponents. 'Compatibility' is relevant during self-play (in self-play the opponent player applies the same algorithm.). 'Safety' ensures that the agent is able to perform in a robust manner against all other algorithms. Desirable system properties to be maintained in Parkes et al. (1997) are efficient coordination and robustness to manipulative behavior. Busoniu, et al. (2010) categorized the suggested criteria in the literature due to two substantial properties, stability and adaption: 'Stability essentially means the convergence to a stationary policy, whereas adaptation ensures that performance is maintained or improved as the other agents are changing their policies.

Veloso (2002) use the WoLF-PHC in a fully competitive task with promising results even though it does not explicitly model its opponent. Carmel and Markovitch (1996) studied the problem of opponent modelling in game playing. They recursively define a player as a pair of a strategy and an opponent model, which is also a player. Each player acquires a model of the opponents' depth of search by using its past moves as examples. Since the learner has a model of the opponent, it can do better than, for example, Minimax return (see e.g. the adaptation criteria in the precedent footnote). Vidal and Durfee (1998a) studied agents modeling other agents in an information market economy and showed that n-level agents will obtain higher payoffs than other agents in a society full of (n-1)-level agents.[28] For example, a 0-level agent does not do any opponent modeling. A 1-level agent assumes its opponent to be 0-level, 2-level agents model the opponent as being 1-level and so on. Vidal (2010) mentions, however, that computational costs of increasing a modeling level grow exponentially for an agent, whereas the utility gains to an agent grow smaller as other agents in the system increase their modeling level. Indeed, when agents learn fast by means of increasing their modeling levels, the system approaches its equilibrium so the advantages of strategic thinking for players within the system are diminished.

Hu and Wellman (1998) mentioned that modelling other agents can be tricky. Agent draw inferences about learning scenarios of other agents by observing the history of the game, however underlying models of agents are highly sensitive to the agents' assumptions about the opponent's policy. Hu and Wellman (1998) suggest that when there is substantial uncertainty about the level of sophistication of other agents the best policy for creating learning agents is to minimize the assumptions about the other agents' intentions and actions. Durfee (1995) suggests that an agent knows more nested models of other agents it must do much more computation.[29] Durfee (1995) characterizes alternative approaches for

[28] For example in Chang and Kaelbling (2001), a strategic agent *A* (PHC exploiter) might exploit an agent *B* (PHC learner) by modeling the learning method of *B*.

[29] Vidal and Durfee (1995) show that even for the pursuit game the models of interactions can be quite large. The pursuit game pits one player against e.g. three opponents. The opponents pursue the player until caught.

nested modeling in which agents are kept free of deep knowledge including explicit communication, selective searching and hierarchical elaboration of choices.[30] For example, in hierarchical reinforcement learning (Barto and Mahadevan, 2003), one agent pursues its target by setting sub goals and evaluating each action based on whether the corresponded sub goal is achieved.

Following the Durfee's idea (1995), we aim at introducing such an adaptive dynamic pricing model where agents do not model each other explicitly as agents. Basically if modelling other agents explicitly, the learner agents must perform a learning task in two steps: first they draw inferences about learning scenarios of other players by observing the past history of that player and learn some underlying model of opponent's actions. Then, agents solve some optimization problem given the learning scenarios of the opponents. The hierarchical nature of the H-PHC agents expedites the accomplishment of both aforementioned steps without explicitly defining the opponents' model of decision. It helps the agents in the first step because it gives more time to adapt to the opponent player's established beneficial decision hierarchies. The hierarchical approach also improves exploration in the second step because exploration of an optimum solution can take big steps at upper levels of hierarchical abstraction.

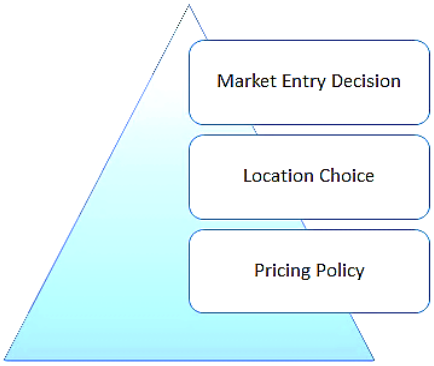
In order to make a sense how real human agents might typically act in the context of strategic interaction in spatial economic models let's think through the Figure 3.2. Assume two dairy processor firms *A* and *B* located in limited area *S* and offering milk delivery contracts to supplying farms distributed in *S*. After some history of firm's interaction assume that both firms follow some pricing policy. At some point, firm *A*'s demand for products at the downstream market rises

[30] Learning based on communication is not a subject of discussion in our paper, although it has been discussed to some extent in the literature. For example in a pursuit game without communication, agents are forced to model each other strictly through observation whereas in Tan (1993) predators can inform each other continuously about their moves. This can give them actual inferences about actual position of prey. Learning based on high-level communication is more complex. Communicative interactions like negotiation and mutual explanation is less studied in the literature (Sian, 1991).

marginally. Consequently, *A* announces it would slightly increase the procurement price of raw milk up to some ε level, since it just wants to encourage more production at farm level. The question here arises is how rational *B* will react to *A*'s policy revision. It is up to firm *B* to now freely revise each dimension of its strategic decision. A rational *B* will probably not directly spoil its established pricing policy in a fully abrupt manner when his price had been advantageous over a long history of firms' interaction. Moreover, firm B must also not necessarily revise its previous strategic decisions for example on choice of location or market exit. It might be more sensible to first invest in minor revisions in his pricing policy or renegotiation of the pricing decision with *A*.[31] Consequently, one might suggest that both players – under non-cooperative price setting or by negotiating – will adjust their price setting through some established hierarchical approach. Solving such a problem in interactive environments is the core methodological issue addressed in this paper.

[31] Apart from the non-cooperative price setting, pricing by negotiating can be potentially the alternative pricing practice in real markets. In fact, cooperative behavior on the part of the processors and joint price setting in agricultural markets has been more extensively studied in the recent literature, especially when it comes to the spatial pricing in raw milk procurement markets (Graubner et al., 2011a; Huber, 2009; Bundeskartellamt 2009; Huber, 2007 a; Huck, Salhofer and Tribl, 2006; Alvarez et al., 2000).

Figure 3.2: Hierarchical decisions in the context of spatially interacting firms.



We restrict our focus in this paper, just to the *pricing policy* of a learning agent setting their market prices in a *non-cooperative* way and show how such an approach can help opposing agents to overcome the described moving target problem with modest computational effort. We introduce a new algorithm, Hierarchical Policy Hill Climbing (H-PHC). In order to evaluate the performance of our algorithm we rely on two desirable properties of Multi-agent learning systems proposed by (Bowling and Veloso, 2002):

Property 1 (Rationality):

> *If the other players' policies converge to stationary policies, then the learning algorithm will converge to a policy that is a best-response to the other players' policies.*

Property 2 (Convergence):

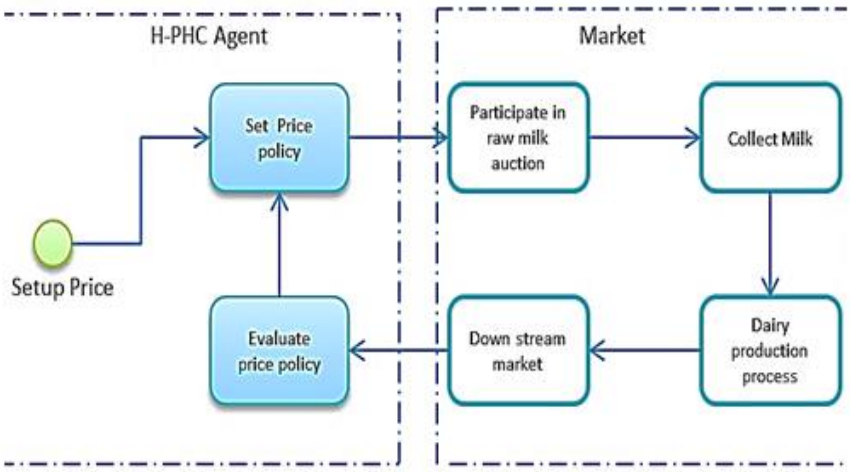
> *The learner will necessarily converge to a stationary policy*.

Property 1 may help agents to keep up with the changes in the environment and constantly move to find optimal behavior. Property 2 is concerned with convergence by concurrent learning entities that are adaptively changing each other's learning environments just in the case there exists Nash equilibria. In line with most studies of MASs we examine our designed agents based on being optimal (in the best-response sense) to the other agents actions and to have a system converging towards the Nash equilibrium. H-PHC algorithm needs to be potentially improved regarding its rationality and can be extended to further hierarchical decision levels.

3.3 The H-PHC algorithm

Figure 3.3 depicts a typical application of the H-PHC algorithm in a spatial raw milk procurement market. The goal of each processor agent is to find the appropriate policy determining the price to bid to spatially dispersed milk farms in order to maximize firm's utility in each period.

Figure 3.3: Agricultural procurement market as simulation environment of H-PHC.



Each H-PHC learner dairy agent has the chance to participate in a raw milk auction for winning raw milk product of spatially dispersed farmers at each round of the game through setting its price level. The supplier farms in our model are price takers, therefore each farmer will choose to deliver its product to the buyer who proposes the highest local bid in the auction stage. When the deals are struck, each buyer agent knows its raw milk suppliers. A pricing policy may be finally evaluated by each H-PHC Agent-based on the gained profit after carrying out the production procedure costs and selling off the dairy products at downstream market. The learner then will set again its revised price in the next round of the game.

The pricing policy in each iteration is evaluated based on a gradient ascent to update agents' policy with regard to pricing hierarchies. A gradient ascent algorithm will start with a suboptimal solution and will improve over time by

changing its policy by small increments. Analogously Agents in H-PHC interaction model preclude non-useful decision hierarchies by means of gradient ascent learning through small increments. This might work as a driver of *rationality* in the agents' behavior. The algorithm might fulfill the *convergent* criterion because once an agent opts out to play a policy -that apparently is beneficial for that agent- it automatically reduces its overall exploration rate (with regard to elaborating new policies). This approach to policy finding by the agent leads to diminished changes in their upper level decision hierarchies giving the opponent more time to learn *rational*. In the following we explain the two major steps of the algorithm depicted in Figure 3.3 i.e. *setting* price policy and *evaluating* price policy.

3.3.1 *Policy setting procedure*

In each round of the game the agent is devoted to conduct its decision across the decision tree - together with further propagation of search tree- from some origin *u* towards most beneficial precise terminal node.[32] Let's assume *P* represents the price of the processor agent at downstream market. In the setup stage, the H-PHC learner agent considers the function $u=unif(0,P)$ as its primitive bidding behavior. *Unif* might describe a random generator function from some continuous uniform distribution of *P, which* generates bids to prospective supplier farms. Indeed, from such primitive initial node onward, the H-PHC agent seeks to generate a sequence of ever-improving solutions by means of an in-depth first search through its decision space. Once the agent has found a solution in the initiation iteration of the game e.g. the primitive bidding behavior , say *u*, the agent will keep improving its search by evaluating *u*'s obtained payoff and deciding upon the direction of the solution space towards which he would like to expand its search. Subject to some exploration policy the agent has to decide in the next iteration whether to continue the developed solution (found in the first iteration) preferably in the direction $u_L=unif(0,P/2)$ or in the direction

[32] Note that the decision parameter is not specified in the original algorithm to be necessarily the price. It could be also quantity or a product of a set of decision variables.

u_R=*unif(P/2,P)*. Partitioning of the solution space might be done broadly by a general halving rule or it can be trickier depending on the nature of the task. Once the agent decides to conduct its search through a parent node (e.g. *u* here) he must invoke for the first time either of the parents' children (e.g. u_L or u_R here). At this point the procedure of search terminates and the result can be revealed as the chosen price of agent in the current iteration.[33] Hence the maximal depth of the search by an agent in the solution space will not be increased further than one step in each period of the game. Further pursuance of an already reached path succeeds in upcoming iterations of the game and depends on Q-values and corresponded probabilities for each child from the primitive solution. Q-values and the corresponding probabilities are the result of policy evaluation in the beforehand iteration (see next section). Indeed the probability of choosing each children node alters after obtaining the payoff after each iteration of the game depending on the policy evaluation procedure. Note that the branching through the decision space might be undertaken subject to some *Cross-Over* probability. In the case of a Cross-Over move the learner will not decompose the searching task into entirely independent sub problems. A parent node might sometimes outsource the task of deepening the search to one of its neighbors. For example assume by branching the node $u_R = unif(p/2, p)$ one agent (subject to some Cross-Over probability) might not only think about its children $u_{RL} = unif(p/2, 3p/4)$ and $u_{RR} = unif(3p/4, p)$ but also it may consider the node $u_{NR} = unif(p, 3p/2)$ as its neighbor. In order to keep away the system from having nodes infinitely assigning credits just to each other, we set 2 constraints on the branching to neighbors. First, neighboring relationship will not lead to reciprocal parent-child relationship. It means no parent node is simultaneously its parent's child. Second, two children e.g. u_{RL} and u_{RR} from the same parent e.g. u_R cannot be simultaneously neighbours of each other. This horizontal delivery of decision choice between distracted vertical paths (*Cross-Over*) is foreseen to examine the caution of

[33] Another termination criterion is when the determined price has achieved a high desired precision from system designer's point of view in terms of its decimal place. This factor in named the *Accuracy* in the algorithm representation. This factor can be given exogenously in setup of the algorithm.

agents' rationality by avoiding overspecializing themselves in upper hierarchical decision levels. Hence more the Cross-Over rate, more the agent is cautious about unilaterally deepening of its search from higher hierarchical decision levels towards terminal nodes. One system designer might propose some effective designing methods for architecture of searching tree by modifying the random determination of adjacent nodes for each node. For example more intelligent agents might be able to revise their decisions by sweeping to some states situated at greater distances within decision tree. In despite of Cross-Over probabilities one can couple some additional mechanism to the algorithm which might serve as further introspection tool for agents. This parameter in our algorithm is *Meta-Decision* and in its humblest form it can impose some minimum (maximum) threshold on probability of undertaking each action despite the learning circumstance. Such small ever ongoing Meta probability will act as a last resort of actions, which has been precluded by the agent previously but need to be animated again. However, like the Cross-Over mechanism, designing such a Meta-Decision mechanism in hierarchical learning can be carried out in different more efficient ways. This is a matter of interest in our ongoing research.

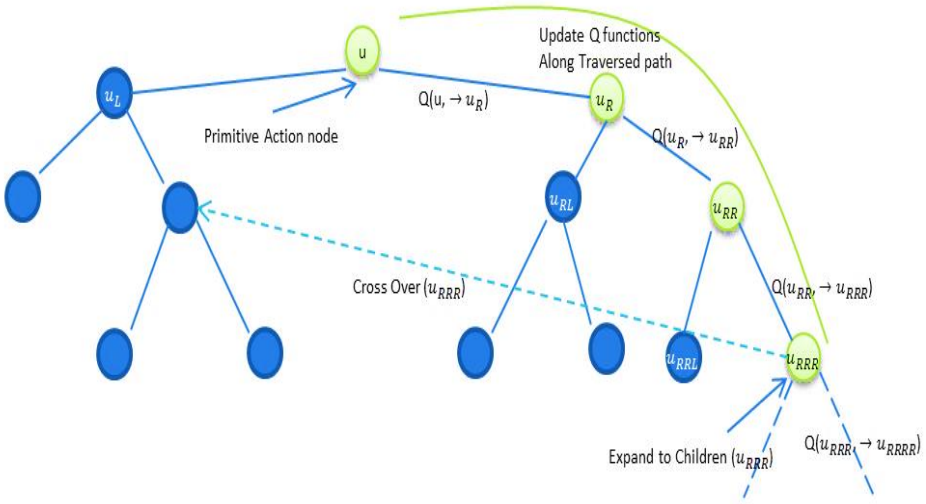
3.3.2 *Policy evaluation procedure*

Assume that in the previous iteration of the learning process the agent pursued some path proceeding from root node (most upper one in its search tree) onward until the terminal node. When the new utility is received (based on feedback of the market), a marker named "Traversed-Path" is attached indicating which nodes the agent has gone through in the decision tree until a modest possible search depth up to the current stage of the game is achieved. In other words during each policy search within one period of the game the algorithm remembers the path traveled across the tree network and returns this path when the search finishes. Our approach to evaluate the usefulness of pursued nodes by selecting a policy is similar to the classic Q-learning (Watkins, 1989). Thereupon the agent assigns its obtained utility in recent round of the game to all nodes embedded in the traversed path and will update the Q-values for each node of traversed path in a

backward direction.[34] Updating a predecessor node along the path is carried out based on the expected reward of its best decision among its successor nodes within the search tree. Hence the Q-value of each decision node is updated through a prediction of the expected utility the agent will receive when conducting its search towards that node and performing the best stored decision following that node. Analogous to Q-learning, in our algorithm some agent specific learning parameters influence the learning procedure of the agent. For example, the Learning-rate *alpha* determines to what extent the newly acquired information will override the old information in each decision node. The factor *Recency* approaching 1 might make the agent counting for only most recent rewards, while factor *Recency* near to 0 renders agent to be some longer term responder to recent obtained rewards. After completion, the updating of Q-values each agent can remember these as a measure for conducting the direction of its decisions with highest Q-values from the Root-state to the terminal nodes. By using a gradient ascent mechanism, the agent modifies its policy just by small increments in favor of the children nodes with the highest Q-value. By using a look up table for probabilities of each children node from the primitive node on, the algorithm can store the new policies (the probabilities of choosing each child from their parent node) at the beginning of next iteration. A typical updating procedure of Q-values is illustrated in Figure 3.4.

[34] By updating the Q-value of a terminal node in traversed path we suppose that the successor of terminate node is the node itself.

Figure 3.4: Policy evaluation procedure between two subsequent rounds of the game.



H-PHC incorporates especially the factor Gradient-Size. Updating the agent's policy by small increments means giving less chance of getting into suboptimal strategies. Agents with higher Gradient-Size parameter will commit sooner to their apparently beneficial decisions. Agents with lower degree of gradient ascent get tied more to their previous decisions more. Hence, and infinitesimally smaller gradient-size might encourage more robust competition within the system. The algorithm is shown in the Figure 3.5 in pseudo code format.[35]

Figure 3.5: H-PHC Algorithm

```
A) SETUP-STEP
SET GRADIENT-SIZE 0<DELTA; 0<ALPHA<1 AND 0<RECENCY<1; ACCURACY /*
PRECISION OF PRICING BY AGENTS*/; 0< CROSS-OVER-RATE<1; 0< META-DECISION-RATE
<1 /* in its simplest form the META-DECISION-RATE imposes some
minimum (maximum) threshold on probability of undertaking
each hierarchy despite the learning circumstance.*/;
TRAVERSED-PATH = EMPTY
```

[35] For access to the code please see online appendix 2 (http://www.ilr.uni-bonn.de/agpo/staff/khalili/khalili.zip)!

```
For each Search direction:
Q(Primitive-action-Node, search-direction)= 0;
 (b) Start-step
Parent-Node ←Primitive-action-Node
Repeat (c) & (d)
(c) Policy-set-step
1.  Reset Traversed-Path
2.  Start-step
3.  Best-found-leaf ←parent-Node
    Add Best-found-leaf Traversed-Path
4.  If Parent-Node Not-Expanded?
    |           |If Size (Parent-Node)> accuracy
    |           |/* SIZE (PARENT-NODE) = Absolute Difference between its limits */
    |           |{Expand-to-children (Parent-Node) Return Parent-Node}
    |           |Else?
    |           {Return Parent-Node}
    ELSE?
                {Parent-Node ←Select-children (Parent-Node)}
                Go to 3.
(d) Policy-eval-step
1.  Go backward through Traversed-Path
    Q(Ancestor Node, search-direction)←Q(Ancestor Node, search-
    direction)+alpha *(Utility + (1-Recency)*
    Max Q(Children, search-direction)- Q(Ancestor Node, search-
    direction))
2. For each (Parent-Node,search-direction):
   |If
   |Q(Parent-Node,search-direction)=MaxQ(Parent-Node,search-direction)
   |   probability ←probability * (1 + delta)}
   |Else?
      probability ←probability * (1 - delta)}
3.  Transform probabilities to the unit interval and adding up
```

```
(E) EXPAND-TO-CHILDREN(PARENT-NODE)
1.  DETERMINE CHILDREN & NEIGHBORHOODS SUBJECT TO CROSS-OVER-RATE
2.  Q-VALUES ←0; PROBABILITIES ←0
(F) SELECT-CHILDREN (PARENT-NODE)
    GAMBLE (CHILDREN) SUBJECT TO META -DECISION /* META- DECISION IS MADE
SIMPLE HERE BUT CAN BE FURTHER ELABORATED*/
```

3.4 **Simulation experiments**

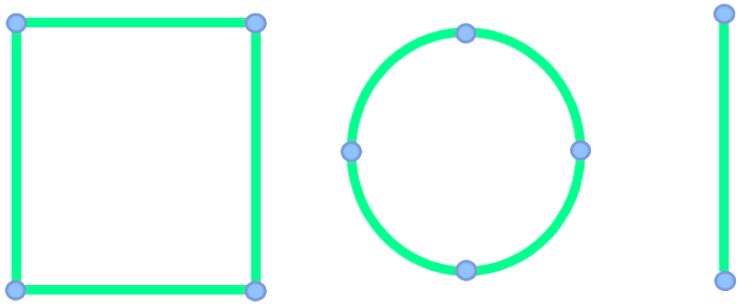
3.4.1 *Markets with free on board pricing*

The prototypical example of spatial price competition is Hotelling's model (Hotelling, 1929). We can adopt this model to our spatial market by considering a finite number of milk supplier farms living uniformly along a linear market. The farms deliver raw milk to two non-cooperative dairy processors (located at each endpoint of the line). An alternative to the linear model in the literature is Salop's circular city model (Salop, 1979). Here multiple processor firms are located on a circle of perimeter 1, equidistant from each other. The supplier farms are uniformly located along the circle. Along with the aforementioned market configurations and without loss of generality we build most of our investigation on an equivalent quadratic market (Figure 3.6), which is a representative of the same equilibria in both Hotelling's and Salop's models and makes the comparison between the simulation results and theoretical findings more convenient.

We first investigate the play between four dairy price-setting processors positioned in the four corners of the quadratic landscape in Figure 3.6. In this experiment we set the term ε (price elasticity of supply) to zero and assume that each farm supplies exactly one unit to one of the buyers. In addition, we assume all sellers (here milk suppliers) receive the same mill price at the buyer's factory gate regardless of their farm locations. Hence, farmers are responsible for costs of transporting the product to the processor (Free On Board or FOB pricing).

Figure 3.6: Configuration of market consisting of spatial milk processors (blue cells) and spatial supplier farms (green cells) simulated as stated by Hotelling's model (right), Salop's model (middle) and Quadratic landscape model (left).

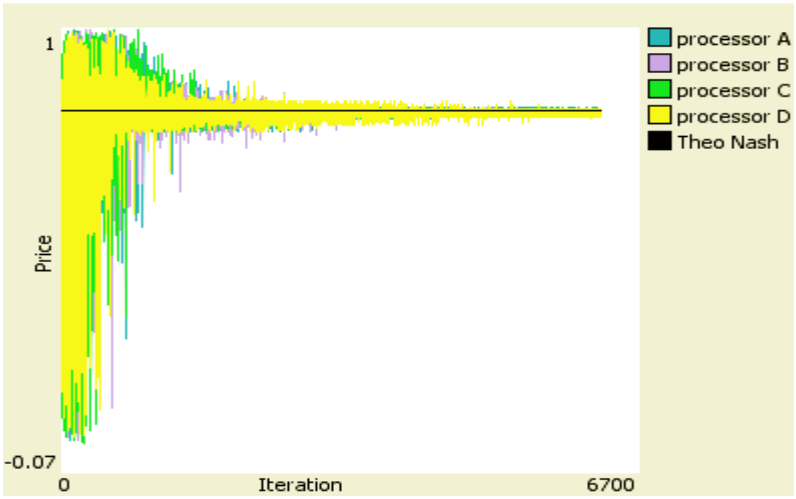
Note: As default we suppose a 20×20 quadratic grid world. The region is discrete in space such that each of green cells can be occupied by one farmer or not. The diameter for each market is assumed to be D=2; hence each distance between two grid cells of the grid world is divided by 10. In addition we limit the values for product price of firms in the downstream market ρ via normalization equal to 1.



The Nash equilibria of all games depicted in Figure 3.6 comprises the price $u^{NE}=\rho$-tD for each agent whereas ρ is the symmetric net product value of processors in down-stream market and t describes a global variable for transportation cost (see Appendix). By designing various experiments, we tested whether H-PHC players successfully converge towards these equilibria through self-play. Figure 3.7 shows an example trajectory of players' strategies while playing the quadratic spatial market game.

Figure 3.7: Price trajectories by simulation of playing spatial market in quadratic landscape.

Note: Transport cost t = 0.1 and $\boldsymbol{u^{NE}}$=.80. Learning parameters of agent are: Accuracy = 0.005, Learning-rate=1, Recency=0.5, delta=0.005, Cross-Over rate=0.1, Meta-decision rate=0.



We presume the parameter t as an explanatory variable of behavior of firms and investigate the outcome of spatial competition between H-PHC agents by exogenously varying the term t. The results of simulating the quadratic market game with regard to a full factorial experimental design of learning parameters are given in table 3.1. We parameterized each agent with some different agent specific values i.e. *Recency* chosen from the list {0.5, 0.75}, *alpha* from {0.9, 1}, *Cross-Over rate* from {0, 0.5}, *delta* from {0.0025, 0.005}. We set the *Accuracy* factor of policy equal to 0.005. Five market types are assumed varying the transport cost chosen from the list {0.1, 0.2, 0.3, 0.8, 1}. Our experiments show that H-PHC learner agents approach the expected Nash equilibrium levels

generally around 5000 iterations (by 4-5% deviation rate) right at earlier iterations of the game (approximately before and about step 5000). At higher transport costs the agents continue for a longer period of time oscillating around the target solution until they cease to change their policy at some point (approximately between step 10000 and 20000). In order to account for all possible volatilities of policies we report the prices in some further steps after the system usually has converged (approximately between step 20000 and 30000).

Table 3.1: Simulations results of learning agents by full factorial experimentation of learning parameters captured in iteration 30000 of the quadratic game and compared with theoretical predictions.

Transport cost rate (t)	**Mean price (standard deviation) of agents**				**Expected Nash equilibria**
	U(A)	U(B)	U(C)	U(D)	U(NE)
0.10	0.802 (0.011)	0.804 (0.010)	0.803 (0.009)	0.805 (0.012)	0.800
0.20	0.591 (0.017)	0.596 (0.012)	0.590 (0.017)	0.589 (0.015)	0.600
0.30	0.389 (0.021)	0.401 (0.028)	0.415 (0.019)	0.411 (0.025)	0.400
0.80	0.490 (0.020)	0.502 (0.004)	0.511 (0.021)	0.501 (0.021)	0.500
1.00	0.510 (0.027)	0.505 (0.018)	0.505 (0.022)	0.510 (0.029)	0.500

The reason for longer searches and some slightly larger standard deviations of prices on markets with higher transport costs has a clear economic reason. Note that the market's competitiveness increases with lower transport cost rates. As the environment gets more competitive, slightly modifying the pricing policy will have larger implications on the agent's payoffs. Consequently, agents have stronger motives for learning more precisely and this then shows in lower standard deviations from the Nash equilibrium price. These motives weaken as transport costs rise.

We further examined the validity of the learning procedure of H-PHC agents by replicating the experiments of Graubner et al. (2011) in order to simulate the

analytical Nash equilibrium from Zhang and Sexton (2001). The former study aims at recapitulating the latter's analytical result using a genetic algorithm approach. The latter study shows that by specifically setting the *price elasticity of supply* equal to 1 and normalizing the dimension of market *D* to 1 there is a symmetric Nash equilibrium comprising the optimal price for each agent in Hotelling's linear model equal to

$$u^{NE} = \frac{2 - 3t + \sqrt{4 + t(13t - 4)}}{4} \tag{3.1}$$

It could be shown that equation 3.1 is a special case of the general equilibrium

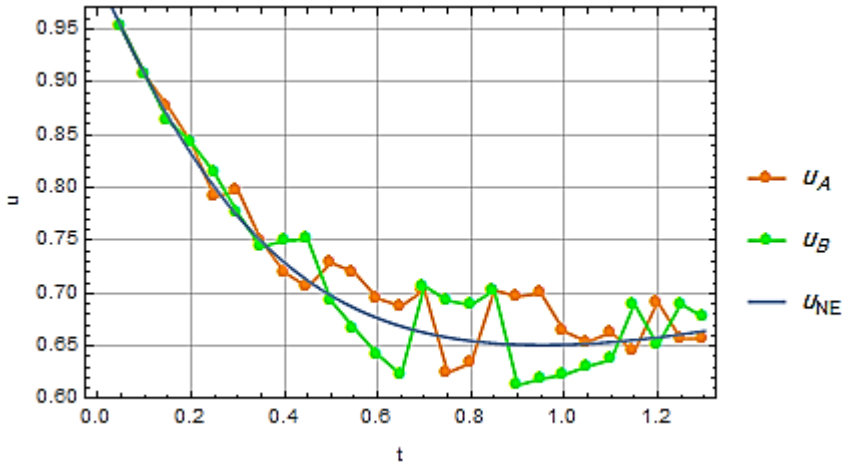
$$u^{NE}(\rho, D) = \frac{2\rho - 3tD + \sqrt{4\rho^2 + t\left(13D^2t - 4D\rho\right)}}{4} \tag{3.2}$$

Equilibrium price 3.2 assumes that $\rho > 3tD/4$.

By systematically changing the transport cost rate $t \in \{0.05, 0.1, \ldots, 1.20, 1, 25\}$, we conducted again Graubner et al.'s simulation runs to test whether the decision making by H-PHC agents is consistent with the expected price behavior above. Similar to our experiments in the quadratic game, H-PHC learner agents in this linear market game narrowed their policy spectrum right at earlier iterations of the game and approach the expected Nash equilibrium level oscillating around the target solution for a relative longer period until they determine their policy at some point. Figure 3.8 represents the optimal policies of agents in iteration 30000 of one random simulation run subject to noticed parameter setting. Please note that changing the learning parameters by switching between the coefficients within the range we used in our examples in this paper or repeating the simulation runs didn't have any substantial influence on the depicted results in Figures 3.8 and 3.9.

Figure 3.8: Comparison between agents' pricing policies in iteration 30000 within one random simulation run and expected theoretical Nash equilibrium in market with 20 discrete farms.

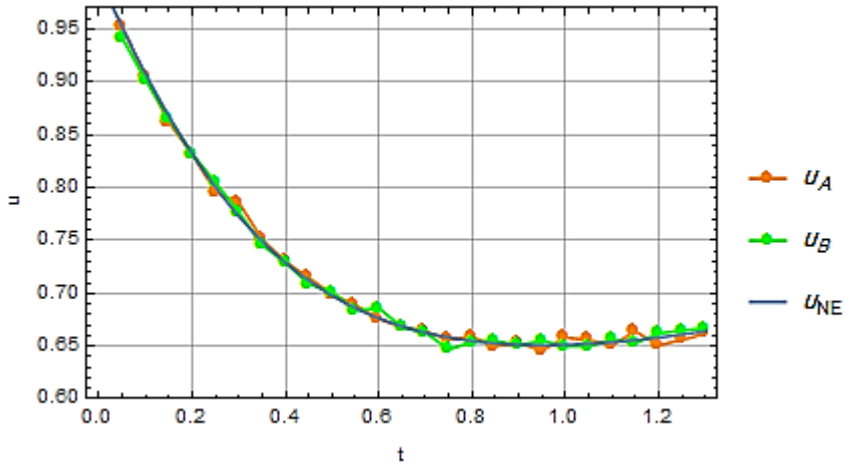
Note: Green circles are presenting processor A and red circles processor B. The correlation coefficients $\boldsymbol{R^2}$ are 0.9176 for agent *A* respectively 0.9218 for agent *B*. Learning parameters: Recency = 0.5, Learning-rate = 1, Cross-Over rate = 0.25, Accuracy = 0.0005, delta= 0.025, Meta-decision rate = 0.



Note that the theoretical model setting in Zhang and Sexton (2001) assumes a continuous number of supplier agents on the line, whereas the simulation environment in our work features just 20 discrete suppliers along a line. The lower number of agents might affect the environment's responses to agents' price setting less smooth and thereby rendering it more burdensome for H-PHC agents to coordinate their pricing policies with higher precision. To more accurately approximate the theoretical equilibrium we increased the number of farms to 200 discrete points on the linear market without changing the length of the market or the learning parameters. We read out the optimal policies of agents at iteration 50000 of experiment. Results show that H-PHC agents are able to very precisely approximate the analytical Nash equilibrium (blue line in Figure 3.9) in the updated market setting.

Figure 3.9: Comparison between agents' pricing policies after iteration 50000 of one random simulation run and expected theoretical Nash equilibrium in market with 200 discrete farms.

Note: Green circles are presenting processor A and Red circles processor B. The correlation coefficients $\boldsymbol{R^2}$ are 0.9972 for agent *A* respectively 0.9976 for agent *B*. Learning parameters: Recency = 0.5, Learning-rate = 1, Cross-over-rate = 0.25, Accuracy = 0.0005, delta= 0.025, Meta-Decision-rate = 0.



3.4.2 *Markets with uniform delivered pricing*

The prevalence of uniform delivery (UD) pricing rules is often observed for agricultural markets. With uniform delivery (UD) pricing farms receive the same price irrespective of their location relative to the processor's production plant. In such market contexts, processor agents are responsible for costs of transporting the product to the processor. The non-existence of pure strategy Nash equilibria in price competition under UD pricing policies -due to discontinuous best response functions of players- is approved in beforehand in the literature of spatial competition (Dasgupta and Maskin, 1986; Beckmann, 1973 and Schuler and Hobbs, 1982). This feature might lead to endless divergences in policy making decisions of players in pure strategies. Cyclic price wars in economies are not only investigated previously in the classic models of pricing in Shubik (1980) but also are discussed in spatial competition models (Schuler and Hobbs, 1982), in MASs (Tesauro and Kephart, 1998) and in evolutionary learning algorithms (Luke and Wiegand, 2002).

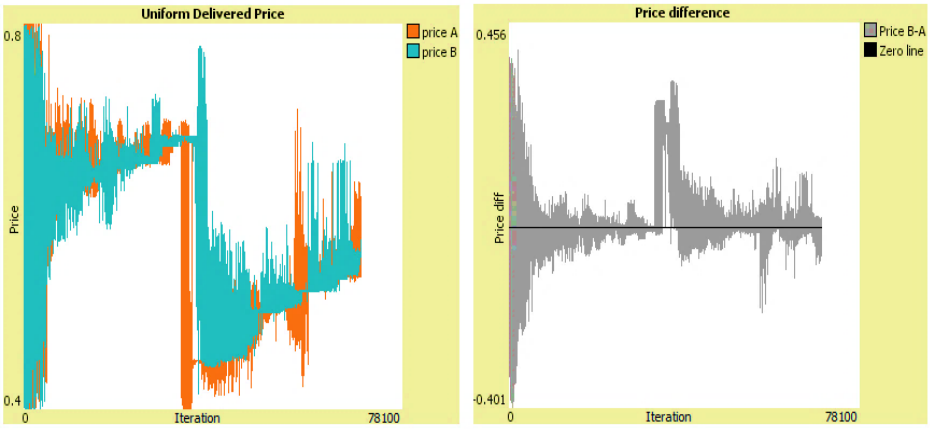
Based on our theoretical foundation, a *rational* learning algorithm is supposed to reflect the instability (cyclic) phenomenon of agents' decisions. Yet without incorporating an appropriate *Meta-Decision* mechanism the expected behavior in H-PHC players' interaction is not achievable. Incorporating the factor *Meta–Decision* in the algorithm can potentially improve this interaction circumstance. The Meta–Decision mechanism supposedly should satisfy the need for "spill over" across actions in game play to encourage experimentation and avoid prolonged fixation on a suboptimal chosen action. Indeed such mechanism is a substantial ingredient of any reinforcement learning model e.g. Roth-Erev model (Roth and Erev, 1995). A Meta–Decision mechanism in *its hierarchical sense* can be complex and is a matter of further deliberation. However, the efficacy of implementing such a factor can be recognized just by adding the most simple trembling hand factor in the algorithm i.e. Meta-Decision.

Figure 3.10 serves to show how imposing some 1 percent ever ongoing minimum threshold on probability of undertaking each hierarchy (despite the learning circumstance) in a typical pricing play between two agents *B* and *A* in Hotelling's linear market (while applying UD pricing) may lead to the anticipated price instabilities. As it can be seen in the Figure, the system begins from a primitive pricing policy applied by both agents and moves toward some convergence in advanced. Hereto agents gradually (with some gradient ascent rate) have eliminated some non-useful decision hierarchies. They stepwise get stuck in narrower and narrower decision levels and are just playing some *semi-equilibrium* play. However the agents' Meta-Decision procedure in active modus can threaten the stability of system at any time. Through random moves triggered by the foreseen Meta-Decision probability, firm *A* will notice that there exists some lower price level that grants a higher utility level for him -compared to the utility of *A* obtained by further playing its semi-equilibrium policy- given the *B*'s price in semi-equilibrium stage of the game. Hence the rational *A* might decrease its price to a deeper level (the first price downswing by agent *A* in Figure 3.10). The recent move of agent *A* renders the current pricing policy of *B* in such high level absolutely nonsense. Since the agent *B* is rational he begins a course of lowering its pricing course by lowering its price to some point marginally above *A*'s price (the reactional delayed first price downswing by *B* in Figure 3.10). However an overbidding cycle is reinitiated as *A* might observe the new situation triggered by

B (recovery of price upswing by agents). Agents continually adapt to each other again and again, never stabilizing at any established price (match). This process is carried out sometimes sporadically depending on agents timing by undertaking their next rational step.

Figure 3.10: Possible non-cooperative interaction of players in an exemplary linear market with complete freight absorption.

Note: Market structure: (t=0.25, ε=1). Learning parameters: Recency = 0.75, Learning-rate = 1, Cross-Over rate = 0.25, Accuracy = 0.0005, delta = .0025, meta-decision-rate = 0.01. For a better declaration, the prices are captured only if agents are not in Meta-Decision (random) moves.



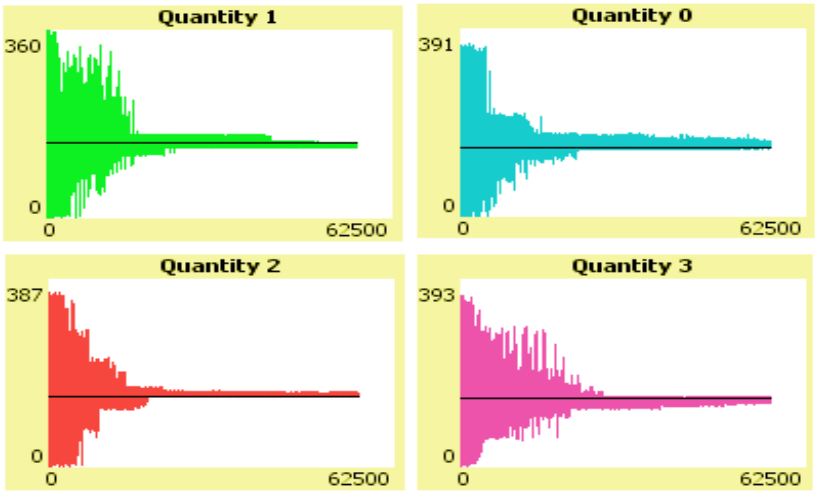
Further simulation experiments show that H-PHC agents incorporating just the simple above mentioned Meta-decision rate cannot guaranty to revive the *rational* property unraveled in the Figure 3.10. Rational agents are supposed to revise their decisions spontaneously if an updating regarding the policies is needed. A dynamic market environment can comprise variety of endogenous shocks like endogenous changing of farms' supply behavior or evolving reaction functions of some firms and etc.. Relaxation of quota regulations in agricultural market in the EU is an example triggering suchlike environmental shocks. Endowing agents with the appropriate Meta-Decision mechanism to guide *when* and *how fast* to get out of the established decisions suited just for preceding market situation is still an open issue to be further deliberated. One concrete fashion of improved implementation of such Meta–Decision mechanism could be learned from the well-known Wolf-PHC (Bowling and Veloso, 2002) algorithm. In this algorithm

by deploying a Meta learning mechanism, the designer imposes on agents to identify whether they are winning or losing. As a consequence of this procedure the agents learn slower when they are winning but faster when they are losing.

3.4.3 *Non-spatial markets*

In the default algorithm we consider the function *u=unif(0,P)* as pricing behavior of buyer agents proposing bids to explicitly model spatial supplier agents. One might imagine domains where each decision node is not allowed to constitute distributive pricing behaviors, for example if the environment is not explicitly modeled or the number of supplier agents is small. In such domains the H-PHC algorithm might work out to evaluate the decision procedure of agents by opting to use the function *u=random(0,P)*. This means that instead of assigning distributed prices to different farms by means of *unif(0,P)*, we assign a random price drawn from the *Random(0,P) to all farms.* We replicate the experiments above by using this price function in the place of the distributive one. The results are qualitatively the same. Agents can converge towards the Nash equilibrium by incorporating the random function as well as the uniform function. This gives evidence that the notion of hierarchical rationality might act as a cognitive motor to encourage non-cooperative coordination in a variety of Multi-agent markets.

Figure 3.11: Players' convergence to expected Nash equilibrium in the Oligopoly system of 4 firms.



In order to examine the feasibility of this idea we conduct our investigation in multiple further domains. As an example, let's study a Cournot Oligopoly system of n firms having the same cost function e.g. $C(x_i) = cx_i$ comprising the exemplary factor *c=0.28* and assuming that the market price is a function of the total supply described by the equation *p=A-aX* with *X* being the total market supply and *A* and *a* being 1 and 0.001, respectively. It can be shown that each firm's optimal quantity in Nash equilibrium amounts is *(A-c)/(a(n+1))* (see appendix). Table 3.2 represents the quantity policies of H-PHC agents in iteration 200000 of 8 simulation runs. Indeed, in the majority of cases equilibrium is achieved between iteration 20000-100000. The more players are involved in the game the more time they need to achieve the coordination. For this we use the same parameters as noted in Figure 3.10 except for the factor *Accuracy* which is set to 1 and *Meta-Decision rate* set to 0. Note that we set the function *u=Random(0,360=Monopol Quantity)* as the primitive bidding behavior for each firm. The total size of interaction space by setting the factor Accuracy to 1 will be 360 power of the players' number.

Table 3.2: Average H-PHC results in the exemplary non-spatial quantity market iteration 200000 of 8 simulation runs.

Number of players	1	2	3	4	5	10
Theoretical Nash	360	240	180	144	120	65.45
H-PHC (Mean & Std-Dev)	(359.8, 0.55)	(239, 3.464)	(179.866, 4.629)	(144.375, 4.541)	(119.55, 4.248)	(65.425, 5.238)

3.5 **Conclusion**

Learning in spatial systems requires algorithms scalable to a large number of agents and that can be implemented with minimal knowledge about the actions of other agents. Most proposed Multi-agent learning algorithms in the literature fail one or both of these criteria. Our research in this paper was set out to develop an operational algorithm in order to lead rational agents to adapt their policy in large-scale and dynamic strategy spaces with modest computational effort. H-PHC learning agents are shown to converge in exemplary spatial games to the best response policies or close to optimal payoffs with minimum required knowledge,

despite the presence of multiple agents in a rich strategic environment. Our experiments show that a community of rational players might be able to overcome the problem of a moving target through hierarchical rationality during the interaction. Knowing more about others' actions causes each agent to reduce the domain of its oscillation regarding its actions. This approach to policy search by a rational agent might let other agents know more about him encouraging the arms race to take place in more confined strategic space. Despite our achievement regarding the efficacy of hierarchical learning, H-PHC in its introduced form needs to be improved to constitute agents capable of *rationally* revise their decisions if an updating regarding the policies is needed. We suggest some additional mechanism named Meta–Decision mechanism to the algorithm to resolve the problem. Precise implementation of Meta–Decision is still an open issue to be further deliberated. One exemplary fashion of Meta–Decision is applied in the well-known Wolf-PHC (Bowling and Veloso, 2002) algorithm.

Explaining the empirically observed collusive behavior (price matching) of spatial agents by defining some higher levels of rationalities is another thinkable path of investigation.

3.6 **References**

Alvarez, A.M., Fidalgo, E.G., Sexton, R.J. and Zhang, M. (2000). Oligopsony Power with Uniform Spatial Pricing: Theory and Application to Milk Processing in Spain. European Review of Agricultural Economics, Vol. 27(3), pp. 347-364.

Barto, A. G. and Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. Special Issue on Reinforcement Learning, Discrete Event Systems Journal, pp. 41-77.

Beckman. (1976). Spatial Price Policies Revisited. Bell Journal of Economics 7, pp. 619-630.

Beckmann, M. J. (1973). Spatial Oligopoly as a Noncooperative Game. International Journal of Game Theory 2, pp. 263–268.

Bellman, R. (1957). Dynamic Programming. Princton, New Jersey: Princton University press.

Bowling, M. and Veloso, M. (2001). Rational and convergent learning in stochastic games. In: Proceedings 17th International Conference on Artificial Intelligence (IJCAI-01), pp. 1021-1026. San Francisco, US.

Bowling, M. and Veloso, M. (2002). Multiagent learning using a variable Learning-rate. Artificial Intelligence 136(2), pp. 215–250.

Busoniu, L., Babuska, R. and De Schutter, B. (2010). Multi-agent reinforcement learning: An overview. In: D. Srinivasan, L. Jain (eds.) Innovations in MASs and Applications - 1, Studies in Computational Intelligence, vol. 310, pp. 183–221. Berlin Heidelberg: Springer.

Camerer, C.F. and Ho, T. (1999): Experience-weighted Attraction Learning in Normal Form Games. Econometrica 67, pp. 827-874.

Carmel, D. and Markovitch, S. (1996). Opponent modeling in MASs. In: G. Weiß,S. Sen (eds.) Adaptation and Learning in MASs, chap. 3, pp. 40–52. Springer Verlag.

Cesa-Bianchi, N and Lugosi, G. Prediction, Learning and Games. Cambridge University Press, 2006.

Chang, Y.H. and Kaelbling, L.P. (2001). Playing is believing: The role of beliefs in Multi-agent learning. In Proceedings of NIPS.

Dasgupta, P. and Maskin, E. (1986). The Existence of Equilibrium in Discontinuous Games. Applications. Review of Economic Studies 53, pp. 27-41.

Durfee, E. H. (1995). Blissful Ignorance: Learning just enough to coordinate well. Proceedings of the first international conference on MASs.

Fudenberg, D. and Levine, D. Theory of Learning in Games. MIT Press, 1998.

Gmytrasiewicz, P. (1992). A Decision-Theoretic Model of coordination and communication in Autonomous systems (Reasoning systems). PhD thesis, University of Michigan.

Graubner, M., Balmann, A. and Sexton, R.J. (2011b). Spatial price discrimination in agricultural product procurement markets: A computational economics approach. American Journal of Agricultural Economics, Vol. 93(4), pp. 949-967.

Graubner, M., Koller, I., Salhofer, K. and Balmann, A. (2011a). Cooperative versus Non-cooperative Spatial Competition for Milk. European Review of Agricultural Economics. European Review of Agricultural Economics, Vol. 38(1), pp. 99-118.

Hotelling, H. (1929). Stability in Competition. Economic Journal, 39 (153), pp. 41–57.

Hu, J. and Wellman, M. (1998). Online learning about other agents in a dynamic multiagent system. In K. P. Sycara and M. Wooldridge, editors, Proceedings of the Second International Conference on Autonomous Agents (Agents'98). pp. 239-246. New York: ACM Press.

Huck, P., Salhofer, K and Tribl, K. (2006). Spatial Competition of Milk Processing Cooperatives in Northern Germany. International Association of Agricultural Economists Conference.

Kash, I, Friedman, E and Halpern, J. (2011). Multi-agent learning in large anonymous games. Journal of Artificial Intelligence Research, 40, pp. 571–598.

Littman, M. L. (1994). Markov games as a framework for Multi-agent reinforcement learning. In Proceedings of the Eleventh International Conference on Machine Learning, pp. 157-163. Morgan Kaufmann.

Löfgren, K. (1986). The Spatial Monopsony:A Theoretical Analysis. Journal of Regional science 26, pp. 7070-730.

Luke, S. and Wiegand, R. P. (2002). Guaranteeing coevolutionary objective measures. See Poli, Rowe, and Jong (2002), pp. 237–251.

Norman, G. (1981). Spatial competition and spatial price discrimination. Review of economic studies, Vol. 48, pp. 345-372.

Parkes, D.C. and Lyle H. U. (1997). Learning and adaption in multiagent systems. Papers from the AAAI Workshop: July 28, 1997, Providence, Rhode Island, ed. S. Sen, 47-52. Menlo Park, C.A.: AAAI Press.

Powers, R. and Shoham, Y. (2005). New criteria and a new algorithm for learning in MASs. Advances in Neural Information Processing Systems, vol. 17, pp. 1089–1096.

Rosin, C. D. and Belew, R. K. (1995). Methods for competitive co-evolution: Finding opponents worth beating. In Stephanie Forrest, editor, Proceedings of the Sixth International Conference on Genetic Algorithms. pp. 373–380. San Mateo,CA: Morgan Kaufman.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. The Bell Journal of Economics Vol. 10, No. 1 , pp. 141-156.

Schuler, R. E. and B. F. Hobbs. (1982). Spatial Price Duopoly under Uniform Delivered Pricing. Journal of Industrial Economics 31, pp. 175–187.

Shoham, Y., Powers, R. and Grenager, T. (2004). On the agenda(s) of research on Multi-agent learning. In Proceedings of Artificial Multiagent Learning. Papers from the 2004 AAAI Fall Symposium. Technical Report FS-04-02.

Shubik, M. (1980). Market structure and behavior. Cambridge, Massachusetts: Harward University Press.

Sian, S. (1991). Extending learning to multiple agents: Issues and a model for Multi-agent machine learning (ma-ml). In Y. Kodrato, editor, Machine learning-EWSL-91 . pp. 440-456. Berlin: Springer-Verlag.

Tan., M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In Proceedings of the Tenth International Conference on Machine Learning, pp. 330-337.

Tesauro, G. and Kephart, J. O. (1998). Pricing in Agent Economies Using Multi-agent Q-learning. Autonomous Agents and MASs, 5, pp. 289–304.

Tuyls, K., Hoen, P.J. and Vanschoenwinkel, B. (2006). An Evolutionary Dynamical Analysis of Multi-agent Learning in Iterated Games. Auton Agent Multi-agent Syst, Volume 12, Issue 1, pp. 115-153.

Vidal, J. M. (1998). Computational Agents That Learn About Agents: Algorithms for Their Design and a Predictive Theory of Their Behavior. PhD thesis, University of Mishigan.

Vidal, J. M. (2010). Fundamentals of Multiagent Systems with Netlogo examples.

Vidal, J. M. and Durfee, E. H. (1995). Recursive agent modeling using limited rationality. Proceedings of the first international conference on MASs. San Francisco: AAAI Press.

Vidal, J. M. and Durfee, E. H. (1998a). Learning nested models in an information economy. Journal of Experimental and Theoretical Artificial Intelligence10(3), pp. 291-308.

Watkins, C. (1989). Learning from Delayed Rewards. PhD Thesis. Cambridge: University of Cambridge.

3.7 **Appendix**

Let's first derive Nash equilibria of the Quadratic game. In order to calculate the total supply obtained by firm *A* (placed at the left upper corner of the quadratic landscape in Figure 3.6) by bidding the price u_A we might identify the supplier farm that is indifferent between delivering its product to either of firms *A* and *B* (placed at the right upper corner of the quadratic landscape in Figure 3.6). Given *D* as diameter of the landscape, the indifferent farm is given by the solution to $u_A - tx_{AB} = u_B - t(D - x_{AB})$. So the supply for firm *A* at the right hand amounts to $X_{AB} = \int_0^{x_{AB}} (u_A - tr)^{\varepsilon} dr$. If ε = 0, we have $X_{AB} = (u_A - u_B + tD)/2t$. Analogously one can calculate the supply for firm *A* at the region underneath of its location in Figure 3.6. By assuming the firm *C* being located at the left lower corner of the quadratic landscape, the total supply of *A* amounts to $X_A = X_{AB} + X_{AC} = (2u_A - u_B - u_C + 2tD)/2t$. Considering ρ_A as the net product value of processor *A* at downstream market and having fixed production

costs C_A, *A* might maximize equation: $\pi_A = (\rho_A - u_A)(2u_A - u_B - u_C + 2tD)/2t - C_A$. Hence at Nash equilibrium of the non-cooperative market, agent *A* will set the price $u_A^{NE} = 0.5\,\{\rho_A - tD + (u_B + u_C)/2\}$. Assuming that other players are delivering their products to the downstream market at the same price and presuming that the other players of the game *C* and *B* are deliberating in the same way as A, leads to the symmetric Nash equilibrium comprising the price u^{NE}=ρ-*tD* for each agent. The derived equilibrium presupposes that the potential market of two competing firms overlap, hence it can be easily shown that this equilibrium price is $\rho > 3tD/2$. By inserting the predetermined parameters of study we conclude that the derived Nash equilibrium price applies if and only if the parameter transport cost is restricted to the interval $0<t<1/3$. Larger transport costs allow the processors to exert monopsony power in isolated markets.

Deriving the Nash equilibrium in a linear Cournot game is as following: Assuming that the price in the market is a function of total supply *X* and calculated as $p=A-aX$ and having the cost function of firm *i* be $C(x_i) = cx_i$, *i* might maximize equation $\pi_i = (A - a(x_i + x_{-i}) - c)x_i$ where x_{-i} represents the sum of supply quantities by all its competitors in the market. Solving this equation for x_i we get $x_i = (A - c - ax_{-i})/2a$. As $x_{-i} = (n-1)x_i$ we obtain $x_i^{NE} = (A - c)/(a(n+1))$.

Chapter 4
A predictive model of pricing by learning agents in spatial agricultural markets

Abstract. Despite some empirical evidence on price formation in spatial agricultural procurement markets, the theoretical explanation of emerging price equilibria is much disputed. The majority of papers attribute the pricing policy of processing firms merely to the spatial structure of markets using a static strategic setting. We analyze the price formation in a dynamic context with a computational approach to overcome analytical limitations in rich strategy space. We show that – in addition to the spatial structure of the market - the pricing behavior of agricultural processors also depends on their ability to learn from each other.

Keywords: learning agents, spatial agricultural markets, ABMs, oligopsony

JEL classification codes: **C63, C72, L13, Q11**

4.1. Introduction

Empirical exploration of agricultural procurement markets often indicate features of imperfect competition (Rogers and Sexton, 1994; Durham et al., 1996; Alvarez et al., 2000; Huck et al., 2006 and Graubner et al., 2011a). Indeed agricultural products are highly perishable goods associated with high storage costs and limited accessibility to alternative buyers. In this market, processors may exercise market power over producers located close to their processing plants. Jointly determined price and quantity decisions can be advantageous to processor firms, but harmful to farmers.

Predicting pricing schemes emerging in spatial markets as equilibria has been a subject of contributions in agricultural economics in recent years. Three pricing regimes are prevalent: Free-on-board (FOB), uniform delivered (UD), and optimal discriminatory (OD) pricing (Beckman, 1976). Until the early 1990s, the researchers' views on the choice of a pricing policy and their market implications are largely based on empirical observations and are quite diverse. Scherer (1980) proposes that the use of UD pricing in industrial markets is associated with a low level of competition. He suggests that in order to increase the degree of competitiveness in markets, firms should not be allowed to price differently from FOB since FOB mill pricing would make the "avoidance of independent pricing more difficult." The view that UD policies are collusive practices is widely established indicating that many industries characterized by high concentration and a spatially differentiated product use UD systems (Zhang and Sexton, 2001). In contrast, Greenhut (1981) and Greenhut et al. (1987) speculate that UD pricing rules emerge in highly competitive markets because it enables firms to compete more effectively over a larger geographical area. Greenhut (1981) investigates the spatial pricing policies of a sample of firms in the United States, West Germany and Japan. According to his observation, the prevalence of spatial pricing policies comprising price discrimination relative to FOB pricing schemes in real world appears to be significant. In addition UD pricing is almost as common in practice as FOB pricing.

Various spatial pricing theories try to capture the emergence of pricing rules and their policy implications in agricultural markets based on characteristics of such markets. Espinoza (1992) and Kats and Thisse (1993) are the first well-known studies theoretically modelling the spatial interaction of processing firms. Both studies suggest that UD pricing systems are likely to be observed in equilibrium for highly monopolistic industries (industries where the transportation cost and/or the discount factor is high) but also in highly competitive industries (industries with low discount factor and transportation cost), while FOB is likely for intermediate market structures. Zhang and Sexton (2001) highlight that the demand function in the studies of Espinoza (1992) and Kats and Thisse (1993) is assumed to be perfectly inelastic leading to bias the firms' choices in favour of UD pricing. Using a supply function with strictly positive (unitary) price elasticity, Zhang and Sexton (2001) suggest that FOB pricing policies emerge as

equilibrium under very competitive structures. Asymmetric FOB–UD regimes are Nash equilibria in less competitive markets and UD pricing emerge when shipping costs are high relative to the value of the finished product, for example for markets that are nearly monopsonistic in nature. Fousekis (2011) adopts Zhang and Sexton's model to specific firm objectives and shows that the co-existence of firms with different objective functions (e.g. profit maximizers and cooperatives) are likely to give rise to some mixed market structures. UD (FOB) pricing is chosen by both competitors in markets where transportation costs are small (large) relative to the net value of the primary product. A mixed FOB–UD pricing equilibrium emerges for an intermediate market structure.

Graubner et al. (2011) apply a computational economics approach using a genetic algorithm. They use in addition a general discriminatory price-based competition approach in two-dimensional space, which is analytically intractable. According to their finding, UD pricing is an equilibrium behavior under relatively mild differentiation between firms (intense spatial competition) and partial but high freight absorption emerges under less intense competition. In contrast to the analytical studies, FOB pricing does not emerge in equilibrium. Table 4.1 depicts a group of representative theoretical contributions regarding spatial pricing forms.

Table 4.1: Theoretical contributions regarding spatial pricing

Work by	Space	Pricing Game	Supply elasticity	Dynamic Model	Specific firm character	Market Equilibrium Outcome
Espinosa, 1992	1-D	Repeated Game	Constant =0	Yes	No	UD pricing in highly monopolistic and highly competitive industries and FOB in intermediate market structures
Zhang and Sexton, 2001	1-D	Static	Constant=1	No	No	FOB in very competitive structures. FOB–UD in less competitive and UD pricing in monopsonistic markets
Fousekis, 2011	1-D	Static	Constant=1	No	IOF or COOP	FOB in noncompetitive structures and mixed FOB–UD equilibrium in intermediate and UD pricing in competitive markets
Graubner et al., 2011	2-D	Repeated Game	Variable	No	No	UD pricing in relatively intense spatial competition and high freight absorption under less competition

The majority of studies investigate pricing policies in spatial markets by static games. An assumption common to above cited studies is that the spatial pricing game is a simultaneous move game and players are assumed to make their decisions in pure or mixed strategies simultaneously in one-shot interactions. Note that just the Spinoza's paper models a reversion to an infinitely repeated static equilibrium right after cheating by any single party. We investigate in this paper a dynamic setting of competition through sequential decisions of agents with foresight.

We presume agents that can anticipate system-wide consequences of their pricing behavior and learn to use a foresight-based decision, rather than following short-run profit maximization rules.[36] Whereas most prior studies in the agricultural

[36] By doing this we can escape a major problem encountered in the predecessor studies of spatial competition namely the nonexistence of pure-strategy Nash equilibria in competitive models (Dasgupta and Maskin, 1986; Beckmann, 1973; Schuler and Hobbs, 1982). This gives rise to conflicting price and output preferences among processor agents. A malicious agent might observe

economics literature attribute the pricing rules by processing firms just to the spatial structure of markets, our research foresees that in addition to the spatial structure of the market - the pricing behavior of agricultural processors also depends on their ability to learn from each other.

The remainder of the paper is organized as follows: After reviewing other relevant literature in section 4.2 and declaring the simulation market context in section 4.3, we introduce details of our foresight based learning mechanism in section 4.4. In section 4.5, we test two opposite poles of processor's learning aptitude i.e. *low-coordination* and *high-coordination* scenarios to draw inferences on the pricing behavior of agents. The final section concludes.

4.2. Computational methods

The study of market power in a broad range of studies in computational economics is often done using genetic algorithms (Vallée and Basar, 1999; Alemdar and Sirakaya, 2003; Arifovic, 1994; Vriend, 2000; Graubner et al. 2011). Agents using a genetic algorithm require less prior competence in the specific task (Arifovic, 1994). Such evolutionary algorithms can be quite useful for some classes of complex problems especially when the problem is non-trivial to deal with. However, *interpreting* the dynamics of genetic algorithms as individual learning processes is not straightforward (Brenner, 2005, p.39). In general, understanding the dynamics of co-evolutionary algorithms is complicated by the fact that the internal fitness measures valuing strategies are *subjective* (Luke and Wiegand, 2002 and Watson and Pollack, 2001). Vriend (2000) and Riechmann (2002) show that the learning dynamics of agents in evolutionary algorithms substantially influences the outcome of the game. For example,

---

the state of its environment including the action of its competitors, and then decides to boost its profit, for example by offering prices above its current price.

assume the vital criterion for measuring the fitness of strategies are relative payoffs compared to a competitor. In this case it pays for an agent to hurt himself (in terms of absolute payoff) as long as he hurts its opponent even more. This type of *spiteful* behavior is a result of the implied algorithm dynamics, not of the game itself.[37]

Given the problems of evolutionary methods, we study the system behavior in a more straightforward way by constituting agents who consciously model their competitors' reactions and incorporate the notion of dynamic programming (Bellman, 1957). A dynamic programming agent foresees upcoming market paths following from his own price setting behavior (Watkins, 1989). Such kind of anticipative learning is also proposed as conjectural variation in the terminology of spatial economics (Capozza and Van Order, 1978).

We investigate how increasing the level of rationality of agents may make a system more robust keeping agents away from exercising persistent price wars. The use of modeling levels as a useful way for agents to classify their knowledge about the world and opponents is established in the literature (e.g. Gmytrasiewicz, 1992 and Vidal, 1998). Basically, heuristic searches regarding modeling other agents have been adopted from the minimax search in games like chess (Carmel and Markovitch, 1996). The minimax principle is applied in a broad literature of artificial intelligence: One agent maximizes her payoff under the worst-case assumption that the opponents will always endeavor to minimize it. In minimax search, an agent finds out her optimal move by exploring the sequence of her actions and her opponents' recursions up to some finite depth of the game. Carmel and Markovitch (1996) study the problem of opponent modelling in game playing. They recursively define a player as a pair of a strategy and an opponent model, which is also a player. Each player acquires a model of the opponents'

[37] In addition one main drawback of evolutionary algorithms mentioned in the literature is their inability to match individual learning histories: agents would remember their past experience to a limited extent. Brenner (2005) states: "it is surprising that the use of genetic algorithms, and especially the original genetic algorithms, has widely spread in simulating economic learning processes."

depth of search by using its past moves as examples. Since the learner has a model of the opponent it can do better than, for example, minimax return. Vidal (1998) studies agents modeling other agents in an information market economy and shows that n-level agents will fare better in a society full of (n-1)-level agents.[38] Vidal (2010) mentioned that computational costs of increasing a modeling level grow exponentially for an agent, whereas the utility gains to an agent grow at a smaller rate as other agents in the system increase their modeling level. Moreover, there would be no adequate system design in such model setting for modeling agents being at the same perception level. By applying the logic of infinitely recursive dynamic programming in our model, we will allow the agents to draw inferences about further progress of the game by looking forward in depth of the future game path without getting into the trap of infinite recursion with opponent models. By doing this, we do not impose any finite constraint for the search depth of agents in learning upcoming stages of the game. Agents in our model are assumed to have an equivalent capability to estimate each other's model. Each player then maximizes a discounted sum of his per period payoffs independent of the history of the game. This method is analogous to the concept of Markov perfect equilibria (Maskin and Tirole, 2001).

4.3. Model setting

We presume two milk processors *A* and *B* located in two-dimensional space represented by a grid of cells and locations are accessible by x-y coordinates. The region is discrete in space such that $X = Y = \{-10, \ldots, 1, 0, 1, \ldots, 10\}$. Following Graubner et al. (2011), a general price equation including all three cases (FOB, UD and OD pricing) is assumed to describe the net price per unit quantity of supply (local price) received by farmers at each location:

[38] A 0-level agent is one that does not recognize the existence of other agents in the world. A 1-level agent recognizes that there are other agents in the world whose actions affect its payoff. A 2-level agent believes that all other agents are 1-level agents. In essence, the n-level agent applies the n-1-level algorithm to all other agents in an effort to predict their action.

$$u_p\left(d_{sp}\right) = m_p - t\alpha_p d_{sp} \tag{4.1}$$

$u_p(d_{sp})$ is the local price of firm $p \in \{A, B\}$ received by supplier farm at point *s*, d_{sp} is the distance between processor *p* and supplier *s*, and *t* describes a global variable for transportation cost rate. The vector (m_p, α_p) is the representative of a processor's pricing policy describing the mill price for a farmer at processor's location by the term m_p and the share of transportation cost absorbed by each farm due to spatial differences of agents expressed by the term α_p. We limit the maximum possible values for product price ρ_p of *A* and *B* in the downstream market via normalization equal to 1. To lower the computational complexity of the system, the price policy parameters of agents m_p and α_p have discrete values between 0 and 1 with predetermined increments of 0.01 and 0.05, respectively. The location of processors is limited to the line between the points (-5,0) and (+5,0). The maximum processors' location distance is normalized to be 1 dividing to 10 (between the points (-5,0) and (+5,0)). Accordingly, each distance between each point of the grid world is also normalized by dividing by 10. We assume that suppliers are price takers and aim for the highest local price offered by processors. The cost function of supplier farms is

$$c\left(q_{sp}\right) = \frac{\varepsilon}{1+\varepsilon}\left(q_{sp}\right)^{\frac{\varepsilon+1}{\varepsilon}} + c \tag{4.2}$$

where $c(q_s)$ is the production cost of producing q_s amount of raw milk and ε is the price elasticity of supply. Following (2) each farm will produce the amount which maximizes its utility function $\Pi^s(q_{sp})$:

$$\begin{aligned} &\max! \leftarrow \Pi^s\left(q_s\right) = u_p\left(d_{sp}\right) q_{sp} - c\left(q_{sp}\right) \\ &\rightarrow q_{sp} = u_p\left(d_{sp}\right)^{\varepsilon} \end{aligned} \tag{4.3}$$

Note that the local prices received by each farm must be positive. Furthermore, the processors do not purchase the raw milk if it does not yield a positive local

profit for them. Hence the set of potential suppliers for each processor *p* is limited within the space by the marginal location at distance r_p:

$$r_p = Min(\frac{m_p}{t\alpha_p} \wedge \alpha_p \neq 0, \frac{\rho_p - m_p}{(1-\alpha_p)t}) \tag{4.4}$$

After submitting the processors' bids to potential suppliers, each processor will earn the local profit Π_{sp} knowing its ultimate supplier calculated as

$$\Pi_{sp} = \rho_p - u_p(d_{sp}) - td_{sp} \quad . \tag{4.5}$$

Ultimately, each processor's utility in our model is the sum of all local profits of its contracted suppliers.

4.4. Simulation design and agent's learning

Assuming the parameter $\rho=1$ for both parties and parameters m_p and α_p discretized into partitions of lengths equal to σ, each agent has a pricing decision space set of length $(1/\sigma + 1)^2$. This triggers $(1/\sigma + 1)^4$ possible pricing interaction between firms. We name each of those interactions a *World-state*. We first divide World-states in 2 categories: *Root-states* and *Non-Root-states*:

*a. Root-state*s are per definition price combinations of *A* and *B* (u_A, u_B) where neither u_A is the best myopic response to u_B nor vice versa is true.

*b. Non-Root-state*s are those states where at least one agent's price is a best myopic response to the other agent's price.

4.4.1. A-level perception

A-level agents are born with a previously acquired knowledge about observing the state of the world and use the best myopic response to boost their utility. The process in which agents will take turns setting prices that are the best myopic response to the opponent is analogous to the process Cournot studied and is expected to have the same long-run property as the simultaneous move adjustment process (Fudenberg and Lewin, 1998, p.11). In order to understand the relevance of A-level-perception we can simulate the market through a series of

distinct repeated interaction stages of the duopsony game. The play begins with the action of one player. In each round of the game, just one agent decides upon price given the price of the opponent. The game then continues in alternating stages.[39] By beginning a dynamic sequential interactive game of A-level agents from any arbitrary history of the game, the system will recognize always a self-enforcing pattern regarding its direction. We name this basic observation induced by observing the behavior of A-level agents the *Basic proposition*.

Basic proposition:

Initiating from an arbitrary starting point (either a Root-state or a Non-Root-state) and driven by sequentially best myopic responses of agents, the system passes always trough a number of Non-Root-states and moves towards some unique set of cyclic World-states (basin of attraction).

We divide Non-Root-states into two categories:

a. *Terminal-state*s consist of the set of Non-Root-states within the basin of attraction.
b. Intermediary-states consist of the set of Non-Root-states which will be linked to the basin of attraction of the game in 1 or more steps from outside of the Terminal-state's cycle.

Figure 4.1: Typical characteristics of Root and Non-Root-states.

Note: Characteristic of Root-state 0: No edges that connect other vertices to 0. Characteristic of Non-Root-states: Number of predecessor and successor node is equal to 1 (1, 4, 2 and 3). Characteristic of Terminal-states: Belonging to the cyclic price basin approachable from an arbitrary Root or Non-Root-state (5 to 10).

[39] Please note that best myopic responses get stored in tables during the simulations through complete search of decision space for each agent and retrieved by repeated use.

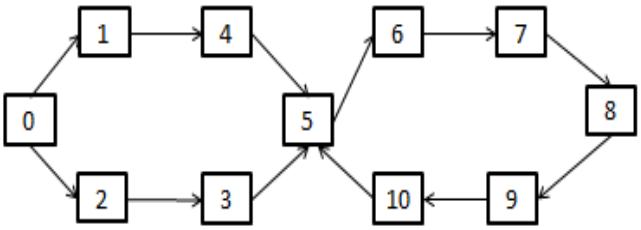


Figure 4.1 illustrates the characteristics of the pricing system if agents pursue their myopic utility starting from an arbitrary state of the world. State 0 represents a typical Root-state where either agent A or agent B can cause the game to move either in direction 0→1 or in direction 0→2. The game will move consequently to the States 2, 3 and 4, which are Intermediary-states and ends up in an infinite circle of terminal price states consisting of states 5-10.

Classic models of price wars, including those introduced by Cournot and Bertrand (Tirole, 1988) have the feature that prices are driven down to a minimum value (e.g. the marginal cost in Bertrand's model). However, limit cycle price wars can potentially arise in markets comprising agents applying freight absorption policies. Such patterns can be explained by attributes of payoff matrices in spatial games due to the well-known nonexistence of pure-strategy Nash equilibria in spatial games (Dasgupta and Maskin, 1986; Beckmann, 1973; Schuler and Hobbs, 1982).

The most basic knowledge integrated into agents' decision in our model is based on backward induction by forward-looking towards *Terminal-states*.

4.4.2. B-level perception

B-level agents incorporate conjectures on opponent's reaction before they set their own prices. Let's assume u_A^* to be the best myopic response of agent A in state S to u_B and u_B^* to be the best myopic response of B in state S to u_A. We define the Deviation-attraction $\varphi_A^{de}(S,i)$ for an agent A in state $S = (u_A, u_B)$ as the long term utility of agent A if agent $i \in \{A,B\}$ decides to act myopically in S foreseeing the payoffs he will get from the next upcoming state S^* onward. Formally:

$$\varphi_A^{de}(S,B)=\begin{cases}\left[\begin{array}{l} if\ u_B \neq u_B^*:\Pi^A\left(u_A,u_B^*\right)+\gamma*\left(Max\left\{\varphi_A^{de}(S^*,A),\varphi_A^{co}(S^*,A)\right\}\right) \\ ,\ S^*=\left(u_A,u_B^*\right)\end{array}\right] \\ if\ u_B=u_B^*:Undifined!\end{cases} \tag{4.6.1}$$

$$\varphi_A^{de}(S,A)=\begin{cases}\left[\begin{array}{l} if\ u_A \neq u_A^*:\Pi^A\left(u_A^*,u_B\right)+\gamma*\left(\varphi_A^{f(B)}(S^*,B)\right) \\ ,\ S^*=\left(u_A^*,u_B\right),\ f(B)=\underset{\omega\in\{de,co\}}{\arg\max}\left\{\varphi_B^{\omega}(S^*,B),\varphi_B^{\omega}(S^*,B)\right\}\end{array}\right] \\ \left[if\ u_A=u_A^*:0\right]\end{cases} \tag{4.6.2}$$

(4.6)

Equation 4.6.1 describes *A*'s Deviation-attraction in the case counterpart *B* undertakes its best response in *S*. The upper bracket implies that *A*'s Deviation-attraction amounts its payoff in state *S** added to *A*'s maximum discounted payoff from *S**. The latter depends on whether it is worth for *A* to cease the best response war in *S** and receive its *Accommodation-attraction* payoff or he foresees that deviation from *S** is worth for him. The lower bracket defines *B*'s deviation non-sense if *B*'s pricing policy in *S* is identical to its best response.

Equation 4.6.2 is describing circumstances, in which *A* decides to deviate from *S*. The upper bracket implies that if the pricing policy of *A* in *S* isn't its best myopic response, he might have benefits (Π^A) from deviation towards state *S**, however he must then account for *B*'s deliberation from *S** whether to accommodate in or deviate from *S**. Hence *A*'s Continuation payoff from *S** then depends on B's decision (deviate or accommodate) in *S**. The lower bracket indicates that if pricing policy of *A* in *S* is its best myopic response, then he has no benefits from deviation from *S* at all.

Analogously we define the Accommodation-attraction $\varphi_A^{co}(S,i)$ and $i \in \{A,B\}$ for an agent *A* in state $S = (u_A, u_B)$ as the persistent utility *A* gains in *S* if both parties agree on price accommodation in *S*:

$$\varphi_A^{co}(S,i)=\frac{Utility\ A\ in\ state\left(u_A,u_B\right)}{1-DiscountFactor} \tag{4.7}$$

Assume *S* is a Non-Root-state of the world. The following knowledge of agents might invite a B-level agent *A* to cease a unilateral price war from *S* onward causing *S* to be a collusive state:

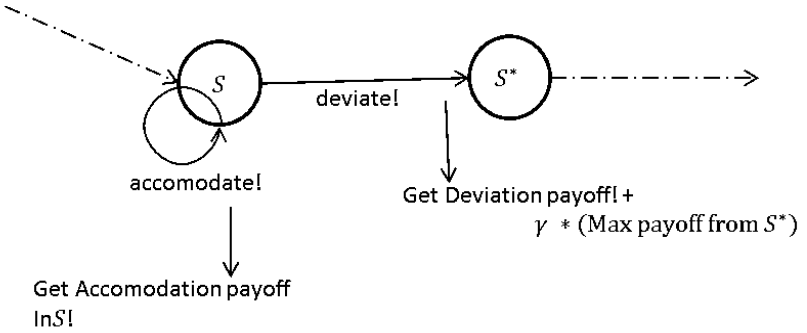
1. Assume $u_B = u_B^*$ and $u_A \neq u_A^*$.

2. *A* knows that $\varphi_A^{co}(S, i) > \varphi_A^{de}(S, A)$ hence will not deviate unilaterally from S.

3. Knowing that *A* will not respond myopically in *S*, *B* also will not respond since (1).

The way of agents' dynamic deliberation by forward looking in each stage of the game can be more simply depicted by figure 4.2.

Figure 4.2: dynamic decision of agents by deviating from or accommodating in state



In line with the mentioned logic in section 4.1, we first implement the above defined functions of Deviation-attractions and Accommodation-attractions of agents in the limited cycle of Terminal-states. In order to overcome the estimation of Deviation-attractions for players in the infinite play of agents in Terminal-states we use a method analogously to the well-known *value iteration* algorithm (Sutton and Andrew, 1998).

The algorithm shown in Figure 4.3 assumes that each agent visits sufficiently[40] all Terminal-states and each time updates the estimation of its long run utility of deviation from one state to the upcoming one. After estimating Deviation-

[40] By using a temporal difference parameter (which will converge to zero), each agent will be able to estimate the attraction of deviation in each state precisely enough.

attractions in Terminal-states and applying Assumptions 1, 2 and 3, each agent put the acquired B-level knowledge in its memory for retrieval in its advanced phases of bilateral price interactions with its competitor.

Figure 4.3: Algorithm 1 for estimating B-level perceptions in Terminal-states[41]

INITIALIZE $\varphi_{\mathrm{I}}^{\mathrm{DE}}(\mathrm{S},\mathrm{I}) = 0$ FOR ALL S FOR AGENT I $\in \{\mathrm{A},\mathrm{B}\}$

LOOP FOR A WHILE T [UNTIL MIN$|| \varphi_{\mathrm{I}}^{\mathrm{DE}}(\mathrm{S},\mathrm{I})^{\mathrm{T}} - \varphi_{\mathrm{I}}^{\mathrm{DE}}(\mathrm{S},\mathrm{I})^{\mathrm{T}-1}|| <$ MIN ERROR, I $\in \{\mathrm{A},\mathrm{B}\}$]:

FOR ALL S: ESTIMATE $\varphi_{\mathrm{I}}^{\mathrm{DE}}(\mathrm{S},\mathrm{I})$, I $\in \{\mathrm{A},\mathrm{B}\}$

4.4.3. C-level perception

B-level agents have learned that optimal behavior of an agent must be conditioned on the expected behaviors of the other agents in the system. For example let's confine our attention just to the Terminal-states in Figure 4.1. Assume state 7 in Figure 4.1 is recognized as a collusive state based on B-level argumentation of agents. Hereby one may argue that agents will change their decision in the system based on the acquired B-level knowledge. Suppose agent *A* would cease the price war in state 7 abstaining from its myopic best response decision based on its B-level knowledge. *B* knowing that *A* has altered its behavior in 7 might e.g. respond differently in state 8, i.e. decides to cease the price war instead of deviating. Knowing that *B* will cease the price war in state 8, *A* might e.g. be deprived of its previous incentive to accommodate in state 7 and so on. It is obvious that agents' recursive modeling can manipulate the strategies-beliefs consistency regarding the behavior of agents in the system just derived through their B-level knowledge. Indeed agents' B-level perception might be self-destructing!

[41] For access to the code please see online appendix 3:
(http://www.ilr.uni-bonn.de/agpo/staff/khalili/khalili.zip)!

The inconsistency problem mentioned before might be healed by designing a mechanism, which alters the behavior of agents until knowing more about the opponent's behavior in upcoming states doesn't manipulate any equilibrium strategies in the system. C-Level perception of agents is then defined as one agents' knowledge which leads to determining agents' decision subject to incorporating the opponents' knowledge about agents' own determined decision. By applying algorithm 2, for example in Terminal-states, we can verify whether collusions obtained in B-level perception would comply with learning termination criteria in C-level perception. Hence, the algorithm 2 assesses the credibility of collusion by reasoning backwards in time through Terminal-states. It proceeds by first assuming that a collusive decision is already compromised by both agents in state S^0. Then it explores what would be the optimal decision in the beforehand state and continues to assess the argumentations along the circle until it again moves to state S^0. If the optimal decision in S^0 is again collusion, then S^0 will be declared as a self-reinforcing C-level collusion.

Figure 4.4: Algorithm 2 for estimating C-level perceptions in Terminal-states

REPEAT UNTIL ALL TERMINAL-STATES OF THE WORLD ARE EXPLORED:

1: CHOOSE ONE OF TERMINAL-STATES S^0 BY RANDOM. ASSUME S^0 INITIALLY TO BE A COLLUSIVE STATE.

2: $S^P \leftarrow$ PREDECESSOR STATE OF S^0.

REPEAT 3 & 4 UNTIL S^0 IS AGAIN THE CHOSEN INITIAL TERMINAL-STATE:

3: ASK THE AGENT X WITH $U_X \neq U_X^*$ IN S^P CHOOSE ITS OPTIMAL DECISION BY BACKWARD INDUCTION FROM S^0.

4: $S^0 \leftarrow S^P$ AND $S^P \leftarrow$ PREDESSECOR S^P; GO TO 3.

5: ASK THE AGENT X WITH $U_X \neq U_X^*$ IN S^0 CHOOSE ITS OPTIMAL DECISION BY BACKWARD INDUCTION FROM SUCCESSOR S^0. IF OPTIMAL DECISION BY X IS ACCOMMODATION, THEN S^0 IS SELF-REINFORCING COLLUSION; ELSE REMOVE S^0 FROM COLLUSIVE STATES.

A summary of the forms of knowledge that agents have or trying to learn in our proposed learning model is depicted in table 4.2.

Table 4.2: Levels of agents perceptionsPERCEPTION LEVEL	ACQUISITION METHOD	LEARNING TERMINATION CRITERIA
A-LEVEL	PREVIOUSLY KNOWN BY AGENTS	MAXIMIZE MYOPIC UTILITY
B-LEVEL	DYNAMIC PROGRAMMING + OPPONENT REACTIONS' OBSERVATION	TEMPORAL ERROR
C-LEVEL	MUTUAL OPPONENT RECURSIVE MODELING	STRATEGY BELIEFS CONSISTENCY THROUGH BACKWARD INDUCTION

4.4.4. Intermediary-states

Agents' knowledge regarding the game's basin of attraction is the basic knowledge of agents' decision *given any arbitrary* history of the game. Until agents know the optimal strategic decision in Terminal-states, the task of agents by reasoning in intermediary respectively Root-states is one of backward induction. For example, assume the path 2→ 3 → 5 in Figure 4.1. Assume in state 3 agent *B* is the firm who is deliberating whether to enter stage 5 of the game or accommodate. In order to perform this task, *B* needs to know about the attraction of its deviation in state 3:

$$\varphi_B^{de}(S3,\mathrm{B}) = \begin{cases} Utility\ B\ in\ \left(U_A,\ U_B^*\right) + DiscountFactor * \varphi_B^{co}(S5,\mathrm{A}), & If\ \ 5 = Collusion\ state \\ Utility\ B\ in\ \left(U_A,\ U_B^*\right) + DiscountFactor * \varphi_B^{de}(S5,\mathrm{A}), & If\ \ 5 \neq Collusion\ state \end{cases}$$

(4.8)

Once *B* finds out its optimal decision in state 3, *A* can learn in the analogous way how to behave in state 2. Respectively agent *B* will acquire the knowledge

regarding his optimal decision in Root-state 0 by having a one-step look ahead to state 2. In order to implement this logical deliberation in the context of simulation, once the values for Deviation-attraction respectively Accommodation-attraction of states in a path are estimated by agents, these will be stored. As soon as the chain of reaction by agents from an arbitrary state of the system forces the system towards encountering previously known states, the continuation pay-offs of agents regarding the remaining path of the game is already stored in memory and will be retrieved by the decision making process of the agents. Algorithm 3 shows the procedure of estimating the Deviation-attraction by agents through backward induction in Intermediary-states.

Figure 4.5: Algorithm 3 for estimating Deviation-attractions and Accommodation-attractions in Intermediary-states

REPEAT UNTIL ALL STATES OF THE WORLD ARE EXPLORED:

1: INITIATE FROM AN ARBITRARY STATE OF THE WORLD

2: OBSERVE THE BEST MYOPIC RESPONSE OF X

3: OBSERVE THE BEST MYOPIC RESPONSE OF THE OPPONENT Y

4: IF NEW STATE OF THE WORLD IS KNOWN CEASE THE EXPLORATION: ESTIMATE OPTIMAL DECISIONS MAKING FOR PREVIOUS STATES OF THE PATH YOU WERE WALKING THROUGH BY BACKWARD INDUCTION; ADD THE WHOLE PATH TO SET OF VISITED STATES; GO TO 1

ELSE: GO TO 2

4.4.5. Root-states

As a general rule, Root-states will be linked with Terminal-states through Intermediary-states. It is imaginable however that there might be some subset of Root-states, which can be linked directly with Terminal-states. For example assume that price combination of *A* and *B* in stages 5-10 in Figure 4.1 are given by the set $\{5\colon (a, b) \rightarrow 6\colon (a^{*}, b) \rightarrow 7\colon (a^{*}, b^{*}) \rightarrow 8\colon (a^{**}, b^{*}) \rightarrow 9\colon (a^{**}, b^{**}) \rightarrow 10\colon (a, b^{**})\}$. Given this set of Terminal-states, one can imagine that, there might be Root-states e.g. (a, b^{*}), (a^{**}, b) or (a^{*}, b^{**}), which display the following characteristic: Agents will face Terminal-states, once one of them foregoes the collusion through unilateral deviation from any of aforementioned Root-states. Attraction values in each Intermediary-state analogously indicate agent's perception regarding continuing the game from that state onward towards Terminal-states. Agents' perception in Root-states likewise incorporates the knowledge of agents in Intermediary-states. In contrast to unilateral collusions mentioned in section 4.2, assembling mutual accommodations by both parties is a pre-condition for establishing collusion in Root-states. The following knowledge of parties would be necessary and sufficient prerequisites to establish *permissible* joint actions in Root-states:

1. *A* knows that $\varphi_A^{co}(S, i) > \varphi_A^{de}(S, A)$
2. *B* knows that $\varphi_B^{co}(S, i) > \varphi_B^{de}(S, B)$
3. *A* knows that *B* knows that (1).
4. *B* knows that *A* knows that (2).

A good formal framework for taking about the knowledge of agents, including the knowledge an agent might have about another agents' knowledge, is given in Vidal (1998) and Fagin et al. (1995). Whether the equilibrium knowledge (or beliefs) of agents emerges based on sharing the knowledge through communication or whether tacit collusions without communication underpin this equilibrium is not included in our model.

4.5. Simulation results

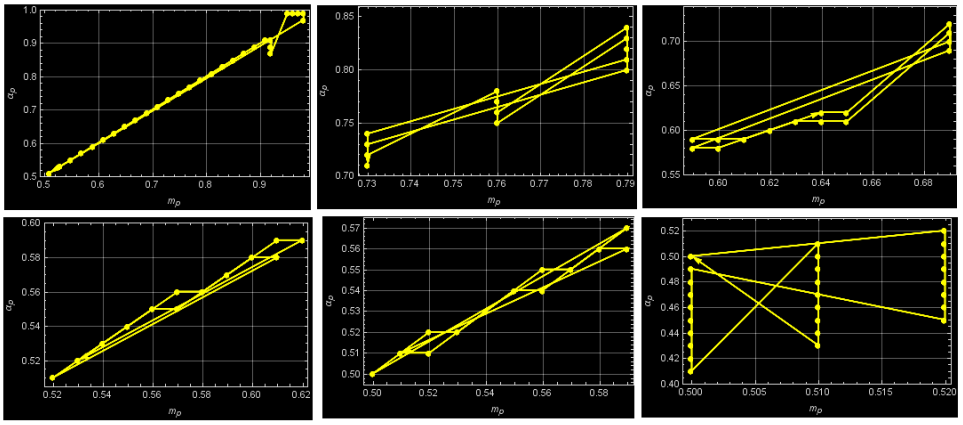
The simulation results cover 2 scenarios, *low coordination* and *high coordination* reflecting 2 opposite poles of our understanding about learning aptitude of processors. The key factor in predecessor studies of spatial competition is the ratio $s=t*D/\rho$ called importance-of-space as an indicator of competitiveness of a market measured by transport costs (t) multiplied distance to competitors (D) divided by net value of product being sold at downstream market (ρ) (Alvarez et al., 2000). As the ratio s increases, competition between firms diminishes to the point where eventually they are spatially isolated monopsonies. Hence, we interpret s as an explanatory variable of our study and investigate the impact of spatial competition by exogenously switching the location of firms toward each other. Simulations are conducted for several selected values of D between 0 (where firms' locations are interlocked) and 1 (maximum processors' location distance). In the first scenario (low coordination), we investigate the flow of the interaction game by A-level players in Terminal-states. In the second scenario (high coordination), we investigate what happens if C-level players learn to cooperate in Root-states.

4.5.1. Low coordination

From the results of our simulations in low coordination scenario, we can derive the following general results in view of the reviewed literature. First, decreasing the factor importance-of-space is triggering cyclic price paths incorporating lower portion of transportation costs absorbed by firms (1-alpha) whereas increasing the Importance-of-space will lead to exercising high freight absorption policies. Second, decreasing the price elasticity of supply will put pressure on farmers' supply price but it also decreases the farmers' freight absorption rate.

Figure 4.6: The processors' cyclic price strategies in Terminal-states based on inter-Firms distances 0 (upper-left hand panel), .2, .4, .6, .8, 1.0 (lower-right hand panel).

Note: Transport cost, supply elasticity and net value of products at downstream market are set to 1. The decision space is discretized to .01 increments. The arrow in each panel serves to follow the direction of pricing sequence.



Figures 4.6 (for the case of elastic supply) and 4.7 (for the case of inelastic supply) show how myopic play will lead to the emergence of a number of non-cooperative cyclic actions in Terminal-states of the game. Each point represents some unstable World-state in the basin of attraction of the game. Such unsteady patterns -suggested in the *basic proposition* - can also be interpreted as market outcomes if agents evaluate their forthcoming payoffs through insignificant discount factors. Note that the set of World-states comprising the basin of attraction of the game is not always unique. For example in addition to the cycles depicted in Figures 4.6 and 4.7 - depending on the previous path of agents'

interaction - further distinct basin of attractions might emerge. We identified concretely all cyclic paths. For the case of the elastic market and *D=0,* two additional terminal cyclic paths are identified by the simulations. One terminal path is indeed encompassing just the unique World-states $(m_p,\ \alpha_p)$ = (0.990, 0.990) for both agents. This point reveals that the system does comprise at least some verifiable *Nash equilibrium* if the locations of firms are interlocked. It is namely recapping the *Bertrand solution* with roughly zero profit for both firms. One possible explanation for such an outcome is that in markets with lower importance-of-space, no party has any incentive to offer freight absorption through UD pricing. The other expected terminal cyclic path in the case of *D=0* in the elastic market includes points that are adjacent to the points depicted in the upper left hand panel of Figure 4.6, thereby resembling a very similar shape.

The emergence of cyclic price competition in markets with low importance-of-space, for example in the upper-left hand panel of Figure 4.6 (D=0), might be understood as a *spatial* case of Bertrand competition in which persistent policy deviations take the role of classic Bertrand Nash equilibrium. In the case of D=0, initiating the price vector $(m_p,\ \alpha_p)$ by (0.500, 0.500) for agent *B* and asking the player *A* to start the sequential A-level perception game, both firms will bid up prices by positively updating both elements of their policy. They travel along the 45 degree line until they achieve a point $(m_p,\ \alpha_p)$, which approximates the Bertrand solution (0.990, 0.990). More precisely, prices travel up to the points (0.980, 0.970) for *A* and (0.980, 0.980) for *B* (see the most upper points in the upper left hand panel of Figure 4.6).

The price competition for *D*=0 is depicted more accurately in Figure 4.8. The utility of firms show a strong oscillating pattern initially, yet in long run they travel along a downward sloping 45 degree line. Both firms continue to bid up reaching the aforementioned prices. The overbidding practice will stop when further overbidding no longer allows any of agents to obtain higher utilities by taking away from the opponent's market area. Contrary to Bertrand competition where prices achieve a maximum value, opponents always preserve some local benefit, i.e. procurement area, in *spatial* competition. Given the last move of *A*, the rational competitor *B* ceases the overbidding procedure and decreases its price again down to the value of its monopsony regime driving the system to the policy points (0.980, 0.970) for *A* and (0.500, 0.500) for *B*. Consequently, agent *A* also

lowers its price to some point marginally above its competitor's monopsony price. Yet the overbidding policy cycle is reinitiated as the beginning situation is now re-launched.

For other ranges of *D* in Figure 4.6 the set of Terminal-states is indeed the unique depicted points. The larger the importance-of-space the larger is the transportation cost absorbed by firms $(1 - \alpha_p)$ approximating the optimal discriminatory pricing point at its maximum (1/(1+1), 1/(1+1)).[42]

Figure 4.7 shows the Terminal-states in the case of having a market with almost non-elastic supply (ε=0.05). Farmers in the reality often have just limited flexibility to substitute outputs creating relatively inelastic supply in the short run (Gardner, 1992).

Figure 4.7: The processors' cyclic price strategies in Terminal-states for different inter-firms distances 0 (upper-left hand panel), .2, .4, .6, .8, 1.0 (lower-right hand panel) with inelastic supply.

Note: Transport cost and net values of products at downstream market are set to 1. The supply elasticity is 0.05. The decision space is discretized to .01 increments. The arrow in each panel serves to follow the direction of pricing sequence.

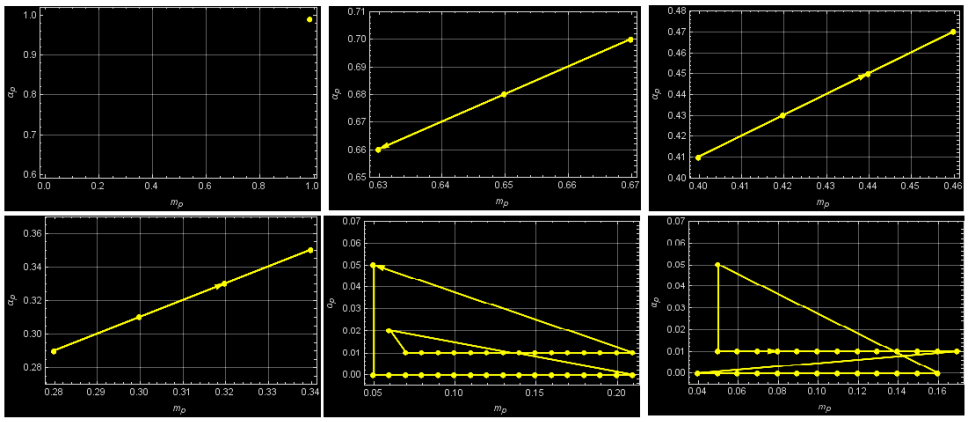


Figure 4.7 shows just one potential cyclic price strategy of firms emerging in Terminal-states based on different inter-firm distances in a typical inelastic

[42] This monopsonistic optimal pricing strategy implies freight cost absorption by the monopsony processor amounts (m_p, α_p) = (1/(1+ε), 1/(ε+1)) (Löfgren 1986).

market. Similar to Figure 4.6, the depicted set of World-states representing the basin of attraction of the game is not always unique. For example, for the case of the *D=0* a total of 2 terminal cyclical paths can be identified. One is demonstrated in Figure 4.7 just including the unique World-states (m_p, α_p) = (0.990, 0.990) for both agents. The other Terminal-state in the case of the *D=0* is the point (m_p, α_p) = (0.990, 1.0) for both agents. Both points confirm that the system setup comprises two Nash equilibria if the locations of firms are interlocked.[43] By setting *D* equal to 0.2, 0.4, 0.6, 0.8, and 1.0 in the inelastic market, the set of Terminal-states is indeed promising a total of 2, 6, 2, 3 and 1 cyclic terminal paths, respectively. The alternative terminal world sets interestingly comprise points that are adjacent to the points depicted in Figures 4.7 resembling some shape very similar to the depicted paths.[44]

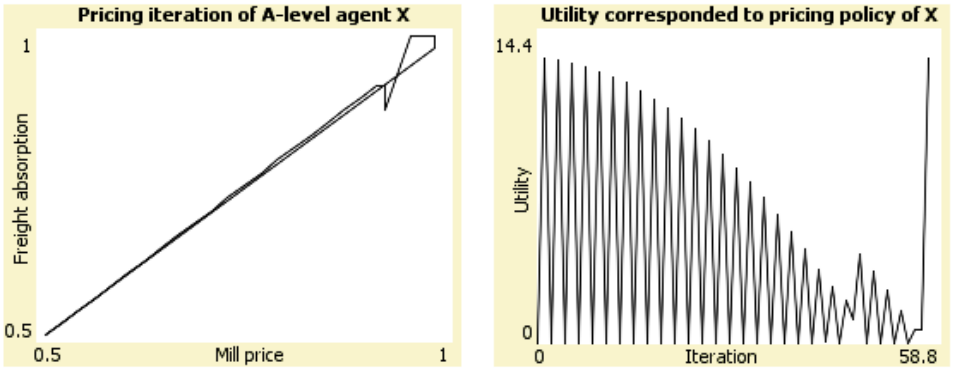
Decreasing the importance-of-space leads to basins of attraction that mimic more FOB pricing wars with the lowest freight absorption emerging in the semi-monopsonistic market (the lower right hand panel of Figure 4.7). Consistent with Figure 4.6, the lowest right hand panel in Figure 4.7 approximates the optimal discriminatory pricing point (.05/(1+.05), .05/(1+.05)).

[43] Our further simulations show that discretizing manner may have an influential on the number of equilibriums but not at the magnitude of the presented freight absorption rates and mill prices.

[44] Pricing policies of other Terminal-states are deviating by 0.01-0.02 increments either in factor m_p or in α_p from each other. As an exception, for *D=.8* in the inelastic market, the other two non-depicted cycles are comprising policies with m_p and α_p numbers around 0.20.

Figure 4.8: Pricing policy path and volatility of utilities by agents applying myopic best response knowledge located according to inter-firm distance = 0

Note: Transport cost, supply elasticity and Net value of products at downstream market are set to 1.



4.5.2. High coordination

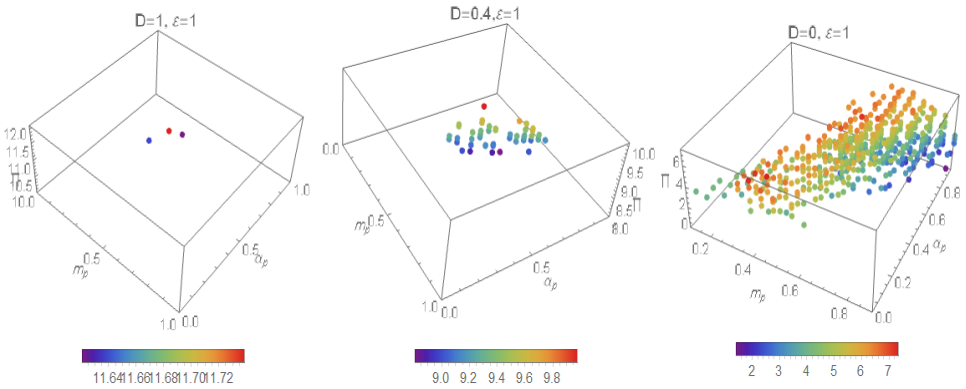
In the reality human agents are capable of looking-ahead to future steps of the game. This might lead to collusive strategic behavior of firms in Root-states dampening the non-cooperative Terminal-state bias. From the results of our simulations in high coordination scenario, we can derive the following general results in view of the reviewed literature. First, if space is less magnificent the range of permissible collusive pricing policies in Root-states can feature some wide scope of mill prices as well as freight absorptions rates deviating from optimal discriminatory price. Second, more situated the firms far from each other at D=1 diverged possible joint policies are assembled approximately around the points depicted by low coordination scenarios!

We computed the set of permissible collusive behaviors of firms in Root-states depending on D for agents having discount factors of 0.25, 0.5 and 0.75. Higher levels of discount factors means that agents take into account future payoffs or they can have deep insight to the compensation possibilities of opponents. The outcome of simulation with lower discount factor levels e.g. 0.25 shows that the agents in majority of market structures might not manage to get out of price disputation in terminal states. However, discount factors of .5 and .75 lead to the emergence of a wide range of permissible cooperative interactions among firms.

Figures 4.9 and 4.10 show collusive pricing policy of firms in elastic respectively inelastic supply market. Note that all obtained permissible pricing policies of firms are not indicating symmetric policies for both firms. It means reaching an implicit consensus among firms does not necessarily comprise identical pricing policies set by firms at equilibrium. Indeed the depicted prices are the set of m_p and α_p of firms that just meet the conditions *1* and *2* in section 4.4.5. However, a significant number of depicted collusive policies are indeed symmetric set equilibrium prices.

Figure 4.9: The processors' equilibrium price policy and corresponded to depicted prices utilities in Root-states based on inter-Firms distances 1, 0.4 and 0 (from left to right) with discount factors 0.75.

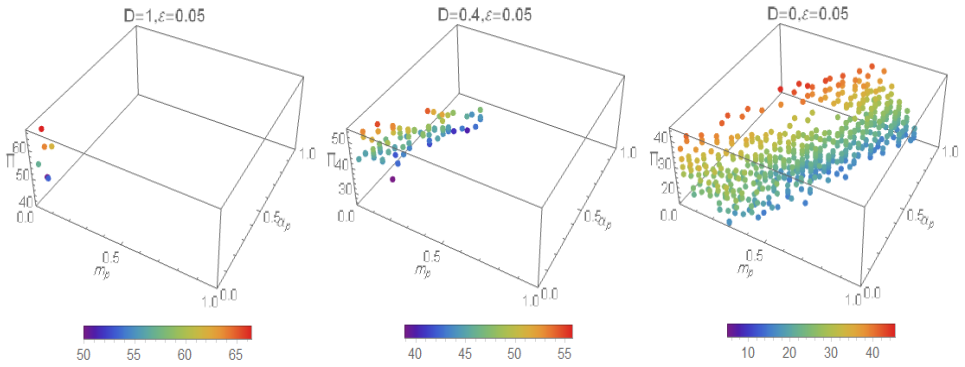
Note: Transport cost, supply elasticity and net value of products at downstream market are set to 1. The decision space is discretized to .05 increments. Colours represent utility ranges.



Figures 4.9 and 4.10 reveal that lowering the factor importance-of-space leads to an increasing number of possible cooperative interactions among firms. This is because by lowering *D* unilaterally actions of agents have stronger impacts on the utility of the counterpart. Hence, in such competitive environments, agents learn more from each other through building price conjectures about the opponent's reactions in upcoming stages of market interaction.

Figure 4.10: The processors' equilibrium price policy and corresponded to depicted prices utilities in Root-states based on inter-Firms distances 1, 0.4 and 0 (from left to right) with discount factors 0.75 and inelastic supply.

Note: Transport cost and Net values of products at downstream market are set to 1. Supply elasticity is .05. The decision space is discretized to .05 increments. Colours represent utility ranges.



The range of permissible collusions between firms in Figure 4.9 and 4.10 get smaller when increasing the importance-of-space from *D=0 to D=0.4*, but still encompasses a variety of freight absorption rates deviating from the optimal discriminatory rates. At *D=1*, collusive policies are assembled approximately around the points $(m_p, \alpha_p) = (0.5, 0.5)$ and $(m_p, \alpha_p) = (0.05/1.05, 0.05/1.05)$ for the case of elastic respectively inelastic supply (resembling distinct monopsony regimes for firms). Consequently, with increasing importance-of-space, learning more about the opponent will not persuade C-level firms to act differently from the low coordination scenario, i.e. for firms with "myopical minds". Hence, advantages of strategic thinking for C-level players within the system are more diminished more the importance-of-space. Summarizing, a larger range of collusive freight absorption rates as well as mill prices are possible with learning agents, but only for lower importance-of-space.

Note that our theoretical model (section 4.4) cannot provide insight in favour of any of its obtained pricing rules. Whether pricing rules granting higher utilities for both firms are implemented with higher probabilities in reality depends on the evolution of knowledge by market players and the possibility to cooperate

between firms. What can be asserted is the existence of some leeway for firms in interactive (non-monopsonistic) market environments for jointly adjusting the elements of their pricing policies e.g. either by jointly offering higher mill prices to farms and simultaneously outsourcing the freight absorption to farms or by offering lower mill prices and being more in charge of transport costs. This violates the assumption that agricultural markets are fully competitive and allows for the hypothesis of abuse of dominant positions in agricultural markets conditional on the factor importance-of-space e.g. the dairy processing sector in Germany (Bundeskartellamt, 2009).

Nevertheless, obtained collusive equilibria in Root-states grant a higher expected utility to processor firms compared to price wars with the difference diminishing as the importance-of-space increases. This remains true no matter whether the collusive pricing policy of firms is symmetric or not. Table 4.3 demonstrates this by comparing expected utilities of agents obtained in Root-state collusions (discount factor = 0.75) with average utilities when iterating in cyclic Terminal-states. In addition the difference between utility of agents *A* and *B* is lower in collusive states (see average utility differences).[45]

Collusions might also lead to smaller volatilities in payoff of firms (see min and max payoffs in Table 4.3). Finally, high coordinative compromises also contribute to the utility maximization intention of the agents. Increasing the inter-firm distance in table 4.3, the utility of firms converges to the utility of OD pricing. As mentioned above already, the increasing the importance-of-space reduces the range of policies to one that assembles around the OD price.

[45] This measure might serve just as an indicator for inequality aversion of firms when establishing collusive strategies.

Table 4.3: The processors' utilities based on foresight (in Root-states) compared to processors' utility through non-cooperative actions (in Terminal-states).

Note: Transport cost, supply elasticity and net value of products at downstream market are set to 1.

	Firms' coordination in Terminal-states (Low coordination)			**Firms' coordination in Root-states (High coordination)**		
Importance-of-space (I)	Min & Max payoff	Expected Payoff	Average Utility difference	Min & Max payoff	Expected Payoff	Average Utility difference
0	0.00, 13.096	3.939	7.803	1.325, 7.363	4.959	1.052
0.2	5.195, 7.040	6.043	0.89	6.118, 8.700	7.437	0.465
0.4	7.740, 9.823	8.669	0.994	8.814, 9.978	9.288	0.279
0.6	8.813, 10.758	9.683	0.755	10.141, 10.990	10.531	0.153
0.8	10.5222, 11.790	11.192	0.648	11.245, 11.448	11.343	0.040
1	11.063, 11.993	11.56	0.448	11.621, 11.739	11.666	0.000

Essentially, we can conclude from table 4.3 that well-coordinated processor agents might seek collusion in order to avoid negative repercussions on their expected utility and price risk. This kind of coordination advantage proposed by our model might be realized by processors in reality, especially if the agents facilitate the coordination through established channels of communication.

4.6. Conclusion

The aim of this paper was to deepen our insight into the spatial pricing in agricultural procurement markets. We investigated the pricing policy of processor

agents in spatial agricultural markets from a game-theoretic perspective. Our interaction scenarios propose that associating the extent of freight absorptions by pricing policy of firms just with factor space in spatial markets might crucially depend on the extent of coordination between firms and hence policy recommendations based on such measures can lead us in the wrong direction.

According to the results of the simulation in markets underlying low coordinative processors scenarios, when competition space is less significant, cyclic equilibrium paths comprising higher partial freight absorption emerge as equilibrium. By increasing the factor importance-of-space price cycles converge eventually to OD pricing. We show that even if price cycles (Terminal-states in our model) mimic the notion of a Nash equilibrium (such that no single agent has any rational myopic incentive to get out of equilibrium), such equilibria are not necessarily efficient outcome for major players in the market, but simply the one that will result from each player individually pursuing his own optimal myopic utility response.

We furthermore reveal that agents who are able to base decisions on what they learn from their future rewards can turn market outcomes around. Our model of agents' coordinative interaction suggests that in addition to the spatial structure of the market, the pricing behavior of agricultural processors also depends on their ability to learn from each other's upcoming reactions. In a world where coordination matters, when competition space is less significant, permissible pricing rules at equilibrium are not bounded by spatial features of the market but encompass a variety of freight absorption rates and or mill prices. By increasing the factor importance-of-space established pricing behavior of firms converge yet again to OD pricing.

4.7. Reference

Alemdar, N. and Sirakaya, S. (2003). On-line Computation of Stackelberg Equilibria with Synchronous Parallel Genetic Algorithms. Journal of Economic Dynamics and Control 27, pp. 1503–1515.

Alvarez, A.M., Fidalgo, E.G., Sexton, R.J. and Zhang, M. (2000). Oligopsony Power with Uniform Spatial Pricing: Theory and Application to Milk Processing in Spain. European Review of Agricultural Economics, Vol. 27(3), pp. 347-364.

Arifovic, J. (1994). Genetic algorithm learning and the cobweb model. Journal of Economic Dynamics and Control 18, pp. 3-28.

Beckmann, M.J. (1976). Spatial Price Policies Revisited. Bell Journal of Economics 7, pp. 619-630.

Beckmann, M.J. (1973). Spatial Oligopoly as a Noncooperative Game. International Journal of Game Theory 2, pp. 263–268.

Bellman, R. (1957). Dynamic Programming. Princton, New Jersey: Princton University press.

Binmore, K. (1987). Modeling rational players, part 1. Economics and Philosophy,3, pp. 179-214.

Binmore, K. (1988). Modeling rational players, part 2. Economic and Philosophy, 4, pp. 9-55.

Brenner, T. (2005). Agent Learning Representation –Advice in Modelling Economic Learning. Germany: Max Planck Institute for Research into Economic Systems.

Capozza, D.R. and Van Order, R. (1978). A Generalized Model of Spatial Competition. American Economic Review, Vol. 68(5), pp. 896-908.

Carmel, D. and Markovitch, S. (1996). Opponent modeling in MASs. In: G. Weiß,S. Sen (eds.) Adaptation and Learning in MASs, chap. 3, pp. 40–52. Springer Verlag.

Dasgupta, P. and Maskin, E. (1986). The Existence of Equilibrium in Discontinuous Games. Review of Economic Studies 53, pp. 27-41.

Durham, C.A., Sexton R.J. and Song J.H. (1996). Spatial Competition, Uniform Pricing and Transportation Efficiency in the California Processing Tomato Industry. American Journal of Agricultural Economics 78, pp. 115-125.

Espinoza, M. P. (1992). Delivered pricing, FOB pricing, and collusion in spatial markets. Rand Journal of Economics 23, pp. 64-85.

Fagin, R. , Halpen ,J.Y. and Vardy M.Y. (1995). Reasoning about knowledge. MIT press.

Fousekis, P. (2011). Free-on-board and Uniform Delivery Pricing Strategies in a Mixed Duopsony. European Review of Agricultural Economics 40, pp. 119-139.

Fousekis, P. (2011). Spatial Price Competition Between Cooperatives Under Hotelling- Smithies Conjectures. Agricultural Economics Review, 12, pp. 4-15.

Fudenberg, D. and Levine, D. K., The theory of learning in games. (1998). Cambridge, MIT Press.

Gmytrasiewicz, P. (1992). A Decision-Theoretic Model of coordination and communication in Autonomous systems (Reasoning systems). PhD thesis, University of Michigan.

Graubner, M., Balmann, A. and Sexton, R.J. (2011b). Spatial Price Discrimination in Agricultural Product Procurement Markets: A Computational Economics Approach. American Journal of Agricultural Economics, Vol. 93(4), pp. 949-967.

Graubner, M., Koller, I., Salhofer, K. and Balmann, A. (2011a). Cooperative versus Non-cooperative Spatial Competition for Milk. European Review of Agricultural Economics. European Review of Agricultural Economics, Vol. 38(1), pp. 99-118.

Greenhut, G. (1981). Spatial pricing in the USA, West Germany and Japan. Economica 48, pp. 79-86.

Huck, P., Salhofer, K. and Tribl C. (2006). Spatial Competition of Milk Processing Cooperatives in Northern Germany. International Association of Agricultural Economists Conference.

Kats, A. and Thisse, J.F. (1993). Spatial oligopolies with uniform delivered pricing. In H. Ohta and J.-F. Thisse (eds), Does Economic Space Matter? New York: St Martins Press, pp. 274-296.

Löfgren, K. (1986). The Spatial Monopsony:A Theoretical Analysis. Journal of Regional science 26, pp. 7070-730.

Maskin, E. and Tirole,J. (2001). Markov perfect equilibrium. Journal of Economic Theory, Vol.20, pp. 191-215.

Sutton, R.S. and Barto, A. (1998). Reinforcement Learning: An Introduction. Cambridge: MIT Press.

Riechmann, T. (2002). Cournot or Walras? Agent-basedlearning, rationality, and long run results in oligopoly games. Discussion paper 261, University of Hannover, Faculty of Economics, K¨onigsworther Platz 1, 30 167 Hannover, Germany.

Rogers, R. T. and Sexton, R.J. (1994). Assessing the Importance of Oligopsony. American Journal of Agricultural Economics 76, pp. 1143–1150.

Luke, S. and Wiegand, R. P. (2002). Guaranteeing coevolutionary objective measures. See Poli, Rowe, and Jong (2002), pp. 237–251.

Scherer, F. M. (1980). Industrial Market Structure and Economic Performance. 2nd edn. Chicago: Rand-McNally College Publishing co.

Schuler, R. E., and Hobbs, B. F. (1982). Spatial Price Duopoly under Uniform Delivered Pricing. Journal of Industrial Economics 31, pp. 175–187.

Sexton, R. J., Kling, C. L. and Carman, H. F. (1991). Market integration, efficiency of arbitrage, and imperfect competition: Methodology and application to U.S. celery. American Journal of Agricultural Economics 73 (3), pp. 568-580.

Shubik, M. (1980). Market structure and behavior. Cambridge, Massachusetts: Harward University Press.

Tirole, J. (1988). The Theory of industrial Organization. MIT Press, Cambridge.

Vallée, T. and Basar, T. (1999). Off-line Computation of Stackelberg Solutions with the Genetic Algorithm. Computational Economics 13, pp. 201–209.

Vidal, J. M. (1998). Computational Agents That Learn About Agents: Algorithms for Their Design and a Predictive Theory of Their Behavior. PhD thesis, University of Mishigan.

Vidal, J. M. (2010). Fundamentals of Multiagent Systems with Netlogo examples.

Vriend, N. J. (2000). An illustration of the essential differences between individual and social learning, and its consequences for computational analyses. Journal of Economic Dynamics & Control 24, pp. 1-19.

Watkins, C. J. (1989). Learning from delayed rewards. Ph.D. thesis, Cambridge University.

Watson, R.A. and Pollack, J.B. (2001). coevolutionary dynamics in a minimal substrate In: Spector, L., et al. (Eds.),. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO). Los Altos: Morgan Kaufmann.

Shoham,Y., Powers, R. and Grenager. T. (2004). On the agenda(s) of research on Multi-agent learning. In Proceedings of Artificial Multiagent Learning. Papers from the 2004 AAAI Fall Symposium. Technical Report FS-04-02.

Zhang, M. and Sexton, R. J. (2001). FOB or Uniform Delivered Prices: Strategic Choice and Welfare Effects. Journal of Industrial Economics 49, pp. 197-221.