Distributed Dynamics and Learning in Games

Bary S. R. Pradelski

Mansfield College
University of Oxford

a thesis submitted for the degree of

Doctor of Philosophy

hilary term 2015

Für Mama und Papa.

Distributed Dynamics and Learning in Games

Bary Sam Raphael Pradelski

Mansfield College, University of Oxford

Doctor of Philosophy in Mathematics

Hilary Term 2015

Supervision:

H. Peyton Young (University of Oxford)

Pierre Tarrès (University of Oxford, Université Paris Dauphine)

Abstract

In this thesis we study decentralized dynamics for non-cooperative and cooperative games. The dynamics are behaviorally motivated and assume that very little information is available about other players' preferences, actions, or payoffs. For example, this is the case in markets where exchanges are frequent and the sheer size of the market hinders participants from learning about others' preferences.

We consider learning dynamics that are based on trial-and-error and aspiration-based heuristics. Players occasionally try to increase their performance given their current payoffs. If successful they stick to the new action, otherwise they revert to their old action. We also study a dynamic model of social influence based on findings in sociology and psychology that people have a propensity to conform to others' behavior irrespective of the payoff consequences.

We analyze the dynamics with a particular focus on two questions: How long does it take to reach equilibrium and what are the stability and welfare properties of the equilibria that the process selects? These questions are at the core of understanding which equilibrium concepts are robust in environments where players have little information about the game and the high rationality assumptions of standard game theory are not very realistic.

Methodologically, this thesis builds on game theoretic techniques and prominent solution concepts such as the Nash equilibrium for non-cooperative games and the core for cooperative games, as well as refinement concepts like stochastic stability. The proofs rely on mathematical techniques from random walk theory and integer programming.

Acknowledgements

I would like to thank my supervisors Peyton Young and Pierre Tarrès. Pierre Tarrès has supported me in developing further mathematical tools, provided detailed comments and has kept asking the right questions to support my work. His abstraction of problems has helped me identify fruitful paths in my research. Peyton Young has been guiding and supporting me since I first became his student during my Masters at Oxford. He has been the supervisor, collaborator, and 'doctorfather' I wished for when reading the "Glass Bead Game" by Hermann Hesse as a teenager. His wisdom, intuition, and understanding have aided me beyond my academic study.

I am indebted to Jörgen Weibull who sparked my fascination for game theory at Ecole Polytechnique and recommended me to Peyton Young. He has been supporting me in my endeavors and is a role-model in his intellectual curiosity and friendship. My visits to Stockholm have been enriching.

Heinrich Nax introduced me to the study of matching markets and we exchanged many ideas and have collaborated. He has been a true friend beyond words. His impact on me is vast, thank you!

I would like to thank Sergiu Hart and Bill Sandholm for detailed comments and being on my thesis committee.

Thanks to Sergiu Hart for hosting me at the Center for the Study of Rationality in Jerusalem, Ehud Kalai for inviting me for a visit to MEDS, Northwestern University, Dirk Helbing and Heinrich Nax for inviting me for a visit to ETH, Zürich, and Garrett van Ryzin and Paul Glasserman for hosting me as a visiting scholar at DRO, Columbia Business School. In all my academic travels I felt very welcome, from undergraduate to senior professor, thank you.

Special thanks for comments and discussions to Itai Arieli, Yakov Babichenko, Peter Biró, Doyne Farmer, Satoru Fujishige, Willemien Kets, Gabriel Kreindler, Jacob Leshno, Jason Marden, Michael Mäs, Ayowande McCunn, Jonathan Newton, Thomas Norman, Max Roser, Johannes Ruf, Jay Sethuraman, Galen Sher and numerous anonymous referees.

I am also thankful for comments by participants at seminars at the 18th CTN Workshop at the University of Warwick, the Center for the Study of Rationality at the Hebrew University, the Technion, the University of Oxford, the IWGTS 2014 in Sao Paulo, the Transatlantic Theory Workshop 2014 in Paris, MEDS, Northwestern University, ETH Zürich, Columbia University, Microsoft Research New England, and the EC 2015 conference in Portland.

I am grateful for support by the Oxford-Man Institute of Quantitative Finance, the Office of Naval Research grant # N00014-09-1-0751, the Air Force Office of Scientific Research grant # FA9550-09-1-0538, and Mansfield College. Special thanks to McKinsey & Company for supporting my educational leave.

My second time at Oxford was as formative as my first. A curious and beautiful period of self-discovery because of good friends and shared moments. Special thanks to my family for their love and support throughout and to Samantha for her love and understanding.

Statement of authorship

The research in this thesis is in parts a result of collaboration with my coauthors:

Chapter 2 is joint work with Peyton Young. The idea and literature review were predominantly conducted by Peyton Young. The model has been joint work and the analysis was predominantly carried out by myself. This chapter is based on the publication:

> Pradelski, B. S. R. and H. P. Young (2012), "Learning efficient Nash equilibria in distributed systems." *Games and Economic Behavior*, 75, 882-897.

Chapter 3 is joint work with Heinrich Nax. Heinrich Nax played the dominant role in the literature review and the setup of the model. The proof of Theorem 12 was a joint effort, while the proofs related to Theorem 14 were predominantly conducted by me. This chapter is based on the publication:

> Nax, H. H. and B. S. R. Pradelski (2014), "Evolutionary dynamics and equitable core selection in assignment games." *International Journal of Game Theory*, forthcoming.

Chapter 4 is independent work. An extended abstract appeared in:

Pradelski, B. S. R. (2015) "Decentralized dynamics and fast convergence in the assignment game.", *Proceedings of 16th ACM Conference on Economics and Computation (EC 15), 2015. Portland, Oregon.*

Chapter 5 is independent work.

Contents

Chapter 1

Motivation

1.1 The big picture

In classical game theory it is commonly assumed that players are perfectly rational. The game is common knowledge and each player is aware of the other players' actions. Every player is also aware that all players act rationally and that every other player is aware of the fact that he acts rationally, ad infinitum. Hence equilibrium play is seen as the logical consequence.

By contrast, *evolutionary game theory* considers boundedly rational players who – through repeated play – may learn and feel their way to a 'good' strategy (see, for example, Weibull 1995, 1998, Sandholm 2010 for surveys). Evolutionary game theory addresses the long-standing criticism that perfect rationality is unrealistic and its assumption forestalls the analysis of disequilibrium play. As Weibull (1998) points out, this approach has a long history going back to *Malthus, Marshall, Schumpeter* and *Hayek*. Indeed, Nash (1950b) discusses a 'mass interpretation' of the *Nash equilibrium* in his PhD thesis. However, only after the influential work by Maynard Smith and Price (1973) did evolutionary game theory emerge. They introduce the concept of *evolutionary stable* *strategy (ESS)* – a strategy which can not be invaded by any other strategy. However, although ESS provides a stability criterion, it is essentially a stationary concept and does not explain how players arrive at such a strategy (see Weibull 1998).

Therefore researchers went on to investigate behavioral learning rules in order to explain how equilibrium or disequilibrium behavior emerges in a dynamic setting (see Foster and Young 1990, Young 1993a,b, Kandori et al. 1993 for pioneering work and Young 1998, 2004, Fudenberg and Levine 1998, Sandholm 2010 for surveys). Well-known and experimentally tested learning rules are, for example, reinforcement learning and aspiration adjustment theory.

Reinforcement learning was first studied in psychology (Bush and Mosteller 1951, 1955, Roth and Erev 1995, Erev and Roth 1998) which in turn follows Thorndike (1898)'s classic model of trial and error: "Choices that have led to good outcomes in the past are more likely to be repeated in the future." This model has been extensively studied for some classes of games but little is known about its general mathematical properties (see, for example, Estes 1950, Bush and Mosteller 1951, 1955, Arthur 1991, Roth and Erev 1995, Beggs 2005, Hu et al. 2011).

Aspiration adjustment theory states that 'a subject lowers his aspiration level after a negative impulse, keeps his aspiration level after a neutral impulse, and raises his aspiration level afer a positive impulse' (see, for example, Hoppe 1931, Heckhausen 1955, Sauermann and Selten 1962, Selten 1998). The models developed by Heckhausen (1955) and Sauermann and Selten (1962) have been confirmed in laboratory experiments on bargaining (see Tietz and Weber 1972, Tietz 1975, Weber 1976, Tietz et al. 1978, Scholz et al. 1983). More specifically they find that the "basis of the aspiration levels changes according to the economic situation and is modified by success and failure in the previous negotiation"(Tietz et al. 1978). This assumption has been confirmed recently by Nevo and Erev (2012) and Nax et al. (2013a). We shall adopt it in several of our models.

Note that these behavioral rules are myopic – agents base their behavior on past outcomes and do not try to forecast future developments. Agents play a best response for the immediate future. Alternatively, *smoothed best-response dynamics* suggest that there is no jump between strategies but rather the probability of playing one strategy increases proportionally with its relative payoff advantage (McKelvey and Palfrey 1995).

In recent years the study of learning in games and decentralized dynamics has flourished in economics and beyond. One particularly interesting question is whether play converges to equilibrium when players have little or no information about the behavior of other players. Economists have thus investigated whether such rules exist that converge to Nash equilibrium (see, for example, Hart and Mas-Colell 2003, 2006, Foster and Young 2003, 2006, Young 2009a). This question is mirrored in computer science, and distributed control theory, where the question is asked from a design perspective: can decentralized dynamics be devised such that equilibrium or a desired outcome can be reached? (See, for example, Papadimitriou 2001, Mannor and Shamma 2007, Marden et al. 2014, and Marden and Shamma 2014.)

The aforementioned contributions all focus on non-cooperative game theory. Evolutionary analysis can also be applied to cooperative games, although this has received considerably less attention in the literature. In a cooperative game, players may form coalitions and thus compete to form coalitions. (See von Neumann and Morgenstern 1944 and Nash 1950a for pioneering work.) Many fundamental solution concepts of cooperative game theory are motivated by informal dynamic processes, for example, the *core* (Gillies 1959, Shapley and Shubik 1972) and the *Shapley value* (Shapley 1953). Nash (1953) calls for bridging the gap between cooperative solution concepts and the non-cooperative approach. Harsanyi (1974) articulates the so-called *Nash program* as follows: "[...] we can obtain a clear understanding of the alternative solution concepts proposed for cooperative games and can better identify and evaluate the assumptions to make about the players' bargaining behavior if we reconstruct them as equilibrium points in suitably defined bargaining games, treating the latter formally as non-cooperative games." (See also Serrano 2005 for a review of the Nash program.)

Nevertheless, cooperative game theory has largely been considered outside the domain of evolutionary game theory until quite recently.[1] This program has been termed 'naturalizing the social contract' (Skyrms 1996, Young 1998). It turns out that the learning models for non-cooperative games mentioned above can be extended to the cooperative framework. This allows to give behavioral foundations for cooperative solution concepts and indeed cooperation.

In this thesis we study learning dynamics for both non-cooperative and cooperative games. We are interested in understanding their dynamic behavior, their rate of convergence, the stability of possible equilibria and their welfare properties. Until recently there has been relatively little research on the rate of convergence.[2] Nevertheless, the rate of convergence is a first 'sanity check' as to whether a proposed dynamic can possibly mirror human behavior in large populations. If the expected time to convergence scales exponentially in the number of players in the game, convergence is so slow that it is unreasonable to expect that it is predictive of actual behavior. Therefore, we believe that

[1] See, for example, Roth and Vande Vate (1990), Agastya (1997, 1999), Arnold and Schwalbe (2002), Newton (2010, 2012), Chen et al. (2011), Rozen (2013), Nax and Pradelski (2014), Newton and Sawa (2015).

[2] Examples for non-cooperative game theory include Ellison (1993), Young (1998, 2011), Morris (2000), Montaneri and Saberi (2010), Kreindler and Young (2013). For cooperative game theory see, for example, Bayati et al. (2005, 2014), Salez and Shah (2009), Celis et al. (2010), Ackermann et al. (2011).

its study is of central importance in evolutionary game theory.

Methodologically this thesis studies the classical equilibrium concepts of game theory. For non-cooperative game theory the most prominent concept is the Nash equilibrium – a state in which any unilateral deviation is not beneficial (Nash 1950b). For cooperative game theory several concepts exist. We focus on the core and its refinements (Gillies 1959, Schmeidler 1969, Shapley and Shubik 1972, Maschler et al. 1979). The core consists of states in which no coalition of players exists such that if the players in the coalition decide to deviate there is a state they can deviate to such that every player in the coalition is at least as well off and at least one player is better off. In order to understand the dynamic behavior of the processes we make use of different branches of mathematics such as Markov, large deviation, and graph theory, random walks and integer programming.

The remainder of this chapter introduces two key concepts that will be used throughout this thesis. *Completely uncoupled learning* describes a class of learning models where players have no information about the actions or payoffs of other players. *Stochastic stability* is a concept that identifies the most probable states in a stochastic evolutionary setting.

1.2 Completely uncoupled learning

Traditionally, economists assumed perfect rationality and knowledge of the other players' actions and payoffs. Foster and Vohra (1997) and Hart and Mas-Colell (2000, 2003, 2006) introduced the class of *uncoupled learning rules* where a player has no knowledge of other players' payoffs (but may know other players' actions).[3] In a *completely uncoupled* (Foster and Young 2006)[4] learning rule players base their behavior only on *their own realized payoff* (and not on other players' payoffs or actions). Such adaptive procedures have been studied before in psychology (Bush and Mosteller 1951, 1955) and more recently in game theory (Erev and Roth 1998, Karandikar et al. 1998). Completely uncoupled dynamics have two advantages:

First, they provide a more realistic model of economic behavior in situations where little information is available to the players or they may not even be aware to be part of a 'game'. Note that even if (some) more information is available or players are (somewhat) more sophisticated, completely uncoupled learning can be understood as one of the two extreme points (perfect rationality being the other) of the possible space of realistic human behavior.[5]

Second, given the absence of information about other players' actions or payoffs, completely uncoupled dynamics allow to avoid the treatment of questions arising from strategic agents. This is justified by the mere fact that there is no useful information to base strategic play on. Thus we can model the interactive dynamics of play without specifying players' strategies.

Our models in chapters 2, 3, and 4 all consider completely uncoupled learning rules. Only chapter 5 studies a learning rule where information about other players actions is directly taken into account, thus an uncoupled but not completely uncoupled learning rule.

[3]Hart and Mas-Colell (2003) show that, in general, such dynamics do not lead to equilibrium play. See Hart and Mas-Colell (2013) for a self-contained collection of articles on the topic.

[4]This concept was first introduced in Foster and Young (2006). We adapt the terminology of Young (2009b).

[5]To date, there is no work on how this space may be axiomatized. A model which, for example, fits in between these extreme points is level-k learning (see Nagel 1995 and Stahl and Wilson 1995).

1.3 Markov processes and stochastic stability

We shall first recall the standard definitions and results for Markov processes and then define perturbed Markov processes and stochastic stability.

1.3.1 Markov processes

A *discrete-time, time-homogeneous Markov process* P is a stochastic process on a finite state space Ω in discrete time steps, $t = 0, 1, 2, \ldots$. The process starts in time $t = 0$ in some state $Z^0 \in \Omega$, determined by a given probability distribution over the state space. The process successively moves from one state to another (or remains in the current state) governed by time homogeneous transition probabilities. The state at time t is denoted by Z^t. We shall write $p_{ZZ'} = \mathbb{P}[Z, Z']$ for the probability of moving from Z to Z' in one time step for $Z, Z' \in \Omega$. Let $P = (p_{ZZ'})_{Z,Z' \in \Omega}$ be the transition matrix. Denote the probability of moving from Z to Z' in $\kappa \in \mathbb{N}$ steps by $p^{\kappa}_{ZZ'} = \mathbb{P}^{\kappa}[Z, Z']$. Accordingly let $P^{\kappa} = (p^{\kappa}_{ZZ'})_{Z,Z' \in \Omega}$.

A state Z' is called *accessible* from Z if there exists a $\kappa > 0$ such that $p^{\kappa}_{Z,Z'} > 0$. Z and Z' *communicate* if Z' is accessible from Z and vice versa. Note that *communication* is an equivalence relation, partitioning the space into *communication classes.* A *recurrent class* is a communication class, such that no state outside the class is accessible from any state inside. Note that it can be shown that every finite Markov process has at least one recurrent class. If the process has exactly one recurrent class, which consists of the whole state space, the process is said to be *irreducible.*

A state Z is *aperiodic* if $P^{\kappa}_{Z,Z} > 0$ for all κ sufficiently large. It holds that if one state is aperiodic then all states are aperiodic (see Norris 1997).

Let $\Pi = (\pi(Z))_{Z \in \Omega}$ be a *stationary distribution* of P if $P(\Pi) = \Pi$ where $\Pi \geq 0$ and $\sum_{Z \in \Omega} \pi(Z) = 1$. We have that:

- There always exists at least one stationary distribution.
- If the process has a unique recurrent class, then Π is unique. The stationary distribution Π describes the time-average asymptotic behavior of the process independently of the initial state Z^0:

  $$\lim_{t \to \infty} \Pi^t(Z|Z^0) = \Pi^\infty(Z|Z^0) = \Pi(Z)$$

  for all $Z \in \Omega$. We say that the process is *ergodic*.
- If the process is ergodic and aperiodic the position of the process at any sufficiently large time is approximated by Π.

Note that we denote a Markov process by its transition matrix P since if the process is ergodic this is sufficient to analyze the long-run behavior of the process.

1.3.2 Perturbed Markov processes and stochastic stability

Let P^0 be a Markov process defined on a finite state space Ω. For each $\varepsilon \in [0, \varepsilon^*]$ for some ε^*, let P^ε be a *regular perturbed Markov process* if P^ε is irreducible for all $\varepsilon > 0$ and $\forall Z, Z'$, $\lim_{\varepsilon \to 0} p^\varepsilon_{ZZ'} = p^0_{ZZ'}$ and if $p^\varepsilon_{ZZ'} > 0$ for some $\varepsilon > 0$, then

$$0 < \lim_{\varepsilon \to 0} \frac{p^\varepsilon_{ZZ'}}{\varepsilon^{r(Z,Z')}} < \infty,$$

for some $r(Z, Z') \geq 0$. For completeness let $r(Z, Z') = \infty$ if $\mathbb{P}^1_\varepsilon[Z, Z'] = 0$. $r(Z, Z')$ is called the *resistance* of the transition from Z to Z'. Hence a transition with resistance r has probability of the order $O(\varepsilon^r)$. We shall call a transition (possibly in multiple periods) $Z \to Z'$ a *least cost transition* if it exhibits the lowest order of resistance, that is, let $Z = Z_0, Z_1, \ldots, Z_\kappa = Z'$ (κ finite) be a path of one-period transitions from Z to Z'. Then a least cost transition minimizes $\sum_{\lambda=0}^{\kappa-1} r(Z_\lambda, Z_{\lambda+1})$ over all such paths.

We are interested in the long-run behavior of a given process when ε is small. We shall employ the concept of *stochastic stability* developed by Foster and Young (1990), Kandori et al. (1993), and Young (1993a). The perturbed process is ergodic for $\varepsilon > 0$ and thus has a unique stationary distribution, say Π^ε over the state space Ω. We are interested in $\lim_{\varepsilon\to 0} \Pi^\varepsilon = \Pi^0$.

Stochastic stability. A state $Z \in \Omega$ is *stochastically stable* if $\Pi^0(Z) > 0$. Denote the set of stochastically stable states by $\mathbf{S}$.

Young (1993a) shows that the computation of the stochastically stable states can be reduced to an analysis of rooted trees on the set of recurrent classes of the unperturbed dynamic. Define the *resistance* between two recurrent classes R and R', $r(R, R')$ to be the sum of resistances of a least cost transition that starts in R and ends in R'. Now identify the recurrent classes with the nodes of a graph. Given a node R, a collection of directed edges T forms a *R-tree* if from every node $R' \neq R$ there exists a unique outgoing edge in T, R has no outgoing edge, and the graph has no cycles.

Stochastic potential. The resistance $r(T)$ of a R-tree T is the sum of the resistances of its edges. The *stochastic potential* of R, $\rho(R)$, is given by

$$\rho(R) = \min\{r(T) : T \text{ is a } R\text{-tree}\}. \tag{1.1}$$

Theorem 1. *(Young 1993a, Theorem 4) Let P^ε be a regular perturbed Markov process, and let Π^ε be the unique stationary distribution of P^ε for each $\varepsilon > 0$. Then $\lim_{\varepsilon\to 0} \Pi^\varepsilon = \Pi^0$ exists and Π^0 is a stationary distribution of P^0 (not* *necessarily unique). The stochastically stable states are precisely these states that are contained in the recurrent class(es) which have minimum stochastic potential.*

In order to identify the stochastically stable states we thus have to follow the general program:

Program 2. *To calculate the stochastically stable states:*

1. *Define the perturbed process and ensure that the conditions hold.*
2. *Identify the recurrent classes of the unperturbed process.*
3. *Calculate the resistance between any two recurrent classes.*
4. *Calculate the stochastic potential for each recurrent class.*

In the following we shall discuss a short example to illustrate the calculation of stochastically stable states. Suppose there are 6 states (dots) as depicted in Figure 1.1. On the left hand side of the figure we show the transition probabilities. Suppose that the residual probabilities are the probabilities to remain in the current state. The right hand side of the figure depicts the recurrent classes. Here arrows are only shown where the resistance of the transition is zero (or equivalently the transition has positive probability in the unperturbed process).

Figure 1.1: Example transition probabilities and recurrent classes.

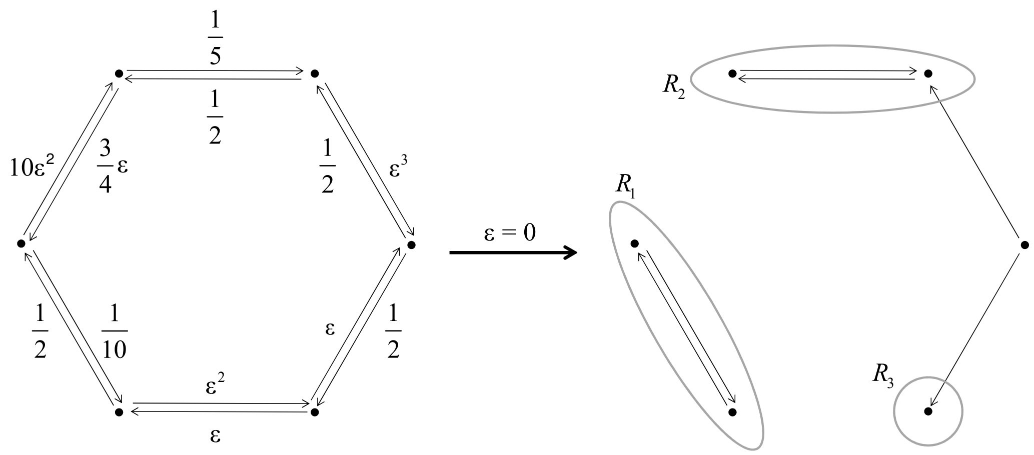
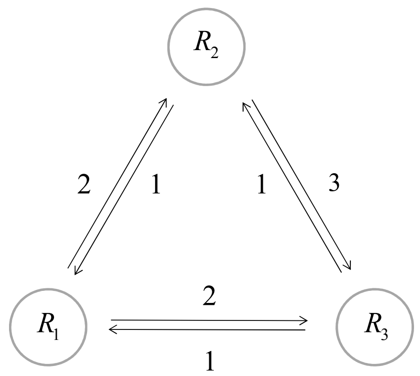


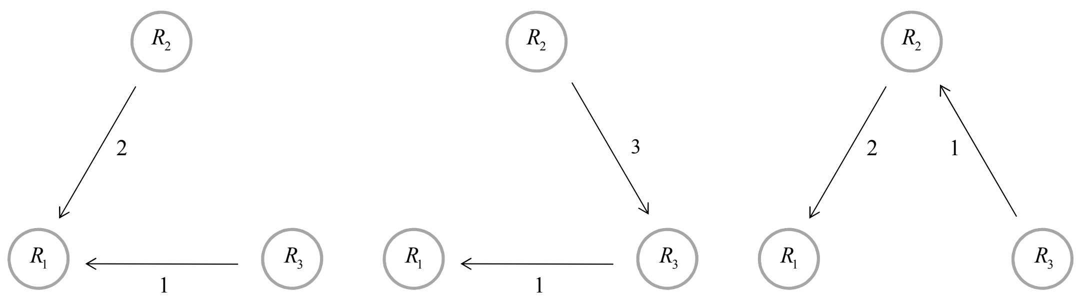
Figure 1.2 shows the resistances between any two recurrent classes.

Figure 1.2: Resistances between recurrent classes.



Now consider for example R_1. The trees rooted at R_1 are shown in Figure 1.3. We find that the minimum among the resistances of the trees is 3 and thus $\rho(R_1) = 3$.

Figure 1.3: Rooted trees for R_1.



In an analogous calculation one finds that $\rho(R_2) = 2$ and $\rho(R_3) = 4$. Now it follows from Theorem 1 that R_2 is the unique stochastically stable recurrent class and thus the two states in R_2 are stochastically stable.

In chapters 2, 3, and 5 we use stochastic stability analysis.

Chapter 2

Learning efficient Nash equilibria in distributed systems

2.1 Learning equilibrium in complex interactive systems

Game theory has traditionally focused on situations that involve a small number of players. In these environments it makes sense to assume that players know the structure of the game and can predict the strategic behavior of their opponents. But there are many situations involving huge numbers of players where these assumptions are not particularly persuasive. Commuters in city traffic are engaged in a game because each person's choice of route affects the driving time of many other drivers, yet it is doubtful that anyone 'knows the game' or fully takes into account the strategies of the other players as is usually posited in game theory. Other examples include decentralized procedures for routing data on the internet, and the design of information sharing protocols for distributed sensors that are attempting to locate a target.

These types of games pose novel and challenging questions. Can such systems equilibrate even though agents are unaware of the strategies and behaviors of most (or perhaps all) of the other agents? What kinds of adaptive learning rules make sense in such environments? How long does it take to reach equilibrium assuming it can be reached at all? And what can be said about the welfare properties of the equilibria that result from particular learning rules?

In the last few years the study of these issues has been developing rapidly among computer scientists and distributed control theorists (Papadimitriou 2001, Roughgarden 2005, Mannor and Shamma 2007, Marden et al. 2009a, Marden et al. 2009b, Asadpour and Saberi 2009 Shah and Shin 2010, Marden and Shamma 2012). Concurrently game theorists have been investigating the question of whether decentralized rules can be devised that converge to Nash equilibrium (or correlated equilibrium) in general n-person games (Hart and Mas-Colell 2003, 2006 Foster and Young 2003, 2006, Young 2009b, Hart and Mansour 2010). A related question is whether decentralized learning procedures can be devised that optimize some overall measure of performance or welfare without necessarily inducing equilibrium (Arieli and Babichenko 2012, Marden et al. 2014). This is particularly relevant to problems of distributed control, where measures of system performance are given (e.g., the total power generated by a windfarm, the speed of data transmission in a communications network), and the aim is to design local response functions for the components that achieve maximum overall performance.

Much of the recent research on these topics has focused on potential games, which arise frequently in applications (Marden et al. 2009b, Marden and Shamma 2012). For this class of games there exist extremely simple and intuitively appealing learning procedures that cause the system to equilibrate from any initial conditions. A notable example is logit learning, in which an agent chooses actions with log probabilities that are a linear function of their payoffs. In this case equilibrium occurs at a local or global maximum of the potential function. However, the potential function need not measure the overall welfare of the agents, hence the equilibrium selected may be quite inefficient. This is a well-known problem in congestion games for example. The problem of inefficient equilibrium selection can be overcome by a congestion pricing scheme, but this requires some type of centralized (or at least not fully decentralized) mechanism for determining the price to charge on each route (Sandholm 2002).

The contribution of this chapter is to demonstrate a simple learning rule that incorporates log linear learning as one component, and that selects an efficient equilibrium in any game with generic payoffs that possesses at least one pure Nash equilibrium. (An equilibrium is efficient if there is no other equilibrium in which someone is better off and no one is worse off.) By 'select' we mean that, starting from arbitrary initial conditions, the process is in an efficient equilibrium in a high proportion of all time periods. Our learning rule is completely uncoupled (see section 1.2). Thus it can be implemented even in environments where players know nothing about the game, or even whether they are in a game. All they do is react to the pattern of recent payoffs, much as in reinforcement learning (though the rule differs in certain key respects from reinforcement learning).

Our notion of selection – in equilibrium a high proportion of the time – is crucial for this result. It is not true that the process converges to equilibrium or even that it converges to equilibrium with high probability. Indeed it can be shown that, for general n-person games, there exist no completely uncoupled rules with finite memory that select a Nash equilibrium in this stronger sense (Babichenko 2012, see also Hart and Mas-Colell 2003, 2006).

The learning rule that we propose has a similar architecture to the trial and error learning procedure of Young (2009b), and is more distantly related to the 'learning by sampling' procedure of Foster and Young (2006) and Germano and Lugosi (2007).[1] An essential feature of these rules is that players have two different search modes: i) deliberate experimentation, which occurs with low probability and leads to a change of strategy only if it results in a higher payoff than the current aspiration level; ii) random search, which leads to a change of strategy that may or may not have a higher payoff. Young (2009b) demonstrates a procedure of this type that selects pure Nash equilibria in games where such equilibria exist and payoffs are generic. However this approach does not provide a basis for discriminating between pure equilibria, nor does it characterize the states that are selected when such equilibria do not exist.

In contrast to these earlier papers, the learning rule described here permits a sharp characterization of the equilibrium and disequilibrium states that are favored in the long run. This results from several key features that distinguish our approach from previous ones, including Young (2009b). First, we do not assume that agents invariably accept the outcome of an experiment even when it results in a strict payoff improvement: acceptance is probabilistic and is merely increasing in the size of the improvement. Second, players accept the outcome of a random search with a probability that is increasing in its realized level of payoff rather than the gain in payoff. Third, the acceptance functions are assumed to have a log linear format as in Blume (1993, 1995). These assumptions define a learning process that selects efficient pure Nash equilibria whenever pure Nash equilibria exist. Moreover when such equilibria do not exist we obtain a precise characterization of the disequilibrium states that have

[1] Another distant relative is the aspiration-based learning model of Karandikar et al. (1998). In this procedure each player has an endogenously generated aspiration level that is based on a smoothed average of his prior payoffs. He changes strategy with positive probability if his current payoff falls below his current aspirations. Unlike the present method, this procedure does not necessarily lead to Nash equilibrium behavior even in 2×2 games.

high probability in the long run. These states represent a trade-off between welfare and stability: the most likely disequilibrium states are those that maximize a linear combination of: i) the total welfare (sum of payoffs) across all agents and ii) the payoff gain that would result from a deviation by some agent, where the first is weighted positively and the second negatively.

2.2 The learning model

We shall first describe the learning rule informally in order to highlight some of its qualitative features. At any given point in time an agent may be searching in one of two ways depending on his internal state or 'mood'. In the content state an agent occasionally experiments with new strategies, and adopts the new strategy with a probability that increases with the associated gain in payoff. (This is the conventional exploration/exploitation form of search.) In the discontent state an agent flails around, trying out randomly chosen strategies every period. The search ends when he spontaneously accepts the strategy he is currently using, where the probability of acceptance is an increasing function of its realized payoff. The key differences between these modes of search are: i) the rate of search (slow for a content agent, fast for a discontent agent); and ii) the probability of accepting the outcome of the search. In the content state the probability of acceptance is determined by the change in payoff, whereas in the discontent state the probability of acceptance is determined by the level of payoff. The rationale for the latter assumption is that a discontent agent will typically try out many different strategies before settling on any one of them. Over time his previous benchmark payoff fades in importance and his current payoff becomes more salient in deciding whether to accept or reject his current strategy.[2] This stands in contrast with the more deliberative behavior of a content agent, who compares the payoff from a single experiment to the benchmark payoff in the immediately preceding period.

A second key feature of the learning process is the mechanism that triggers transitions between content and discontent states. A transition from content (c) to discontent (d) occurs when an agent's realized payoff goes down for two periods in succession and he does not experiment during those periods. In other words, a $c \to d$ transition is triggered by a drop in payoff that was not instigated by the agent, but results from a change of action by someone else. By contrast, a $d \to c$ transition occurs when an agent tires of searching and accepts his current strategy as 'good enough'.

To illustrate these ideas in an everyday situation, consider a commuter who ordinarily takes a particular route to work. A 'content' commuter will occasionally experiment with new routes just to see if there might be something better out there. If one of these experiments results in lower travel time, he adopts the new route with a probability that is increasing with the gain in payoff (the decrease in travel time), and otherwise he returns to his usual route. However, if the travel time on his usual route worsens over several consecutive days, this will prompt him to become discontent and to start actively looking for a new route. This search eventually comes to an end when he adopts the route he is currently using, where the probability of acceptance increases the more desirable the route is.

It is worth pointing out that these acceptance probabilities depend on cardinal changes in payoffs, and therefore go beyond the standard assumptions of von Neumann-Morgenstern utility theory. In particular, the learning model is not

[2]The algorithm could be modified so that an agent compares his current payoff to a discounted version of his old benchmark. We conjecture that this would lead to similar long-run behavior but it would also significantly complicate the analysis.

invariant to changes of scale or changes of origin in the agent's utility function. The larger the payoff gain from an experiment, the higher the probability that it will be accepted; similarly, the larger the current payoff level from a random search, the higher the probability that the current action will be accepted (and the benchmarks adjusted accordingly). Since agents' payoffs determine their probabilities of making various choices, it should not be surprising that the probability of different states of the system depends on their welfare properties. In other words the learning process implicitly makes interpersonal comparisons of welfare among the agents. Our main theorem will exhibit the precise sense in which the long-run probability of the states is related to both their welfare and their stability.

We now describe the learning algorithm in detail. Let $\mathcal{G}$ be an n-person game on a finite joint action space $A = \prod_i A_i$ and let $u_i(\cdot)$ be i's utility function. At each point in time the *state* of each player i is a triple $z_i = (m_i, \bar{a}_i, \bar{u}_i)$ consisting of

1. a mood m_i: a state variable that determines his method of search,
2. a benchmark action $\bar{a}_i$: the action he would take when not searching,
3. a benchmark payoff $\bar{u}_i$: the payoff he expects to get, i.e., his aspiration level.

Each agent has four moods: content (c), discontent (d), watchful (c^-), and hopeful (c^+). We shall assume that each benchmark payoff $\bar{u}_i$ corresponds to a feasible payoff in the game, that is, $\bar{u}_i = u_i(a)$ for some $a \in A$. Since A is finite, it follows that the state space is also finite.

The learning process evolves in discrete time periods $t = 1, 2, 3, \ldots$, where the game is played once per period. The agents' exploration rates are governed by a parameter $\varepsilon \in (0, 1)$ that determines the overall level of *noise* in the system, and two real-valued functions $F(u)$ and $G(\Delta u)$, where $\varepsilon^{F(u)}$ is the probability of accepting the outcome of a discontent search when the current payoff is u, and $\varepsilon^{G(\Delta u)}$ is the probability of accepting the outcome of a content search when the payoff gain is $\Delta u > 0$. We shall assume that F and G are strictly monotone decreasing, that is, the probability of acceptance is strictly increasing in u and Δu respectively. For simplicity of exposition we shall also assume that ε, F and G are common across all agents. In section 2.7 we shall show how to extend our results to the case where agents have heterogeneous learning parameters, including different rates of experimentation.

Fix a particular agent i. Conditional on i being in state z_i at the start of a period, the following rules determine i's state at the end of the period. (The process is time homogeneous, so there is no need to designate the period in question.)

Content: $z_i = (c, \bar{a}_i, \bar{u}_i)$. At the start of the period i experiments with probability ε and does not experiment with probability $1-\varepsilon$. An *experiment* involves playing an action $a_i' \neq \bar{a}_i$, where the choice of a_i' is determined by a uniform random draw from the set $A_i - \{\bar{a}_i\}$. Let u_i' be i's realized payoff during the period whether he experiments or not.

- If i experiments and $u_i' > \bar{u}_i$, i transits to the state $(c, a_i', \bar{u}_i')$ with probability $\varepsilon^{G(u_i'-\bar{u}_i)}$, and remains in state $(c, \bar{a}_i, \bar{u}_i)$ with probability $1-\varepsilon^{G(u_i'-\bar{u}_i)}$.
- If i experiments and $u_i' \leq \bar{u}_i$, i's state remains unchanged.
- If i does not experiment and $u_i' > \bar{u}_i$, i transits to the state $(c^+, \bar{a}_i, \bar{u}_i)$.
- If i does not experiment and $u_i' < \bar{u}_i$, i transits to the state $(c^-, \bar{a}_i, \bar{u}_i)$.
- If i does not experiment and $u_i' = \bar{u}_i$, i's state remains unchanged.

Thus if i experiments and the result is a gain in payoff, he accepts the outcome (and adopts the corresponding new benchmarks) with a probability that increases with the size of the gain. If i does not experiment and his payoff changes due to the actions of others, he enters a transitional state and keeps his previous benchmarks. The notion here is that it would be premature to change benchmarks based on the experience from a single period, but if the directional change in payoff persists for another period, then i transits to a discontent state or a new content state, as described below. (It is straightforward to modify the algorithm so that an agent waits for k periods before transiting to a new content or discontent state.)

Watchful: $z_i = (c^-, \bar{a}_i, \bar{u}_i)$. Agent i plays $\bar{a}_i$ and receives payoff u_i'.

- If $u_i' < \bar{u}_i$, i transits to the state $(d, \bar{a}_i, \bar{u}_i)$.
- If $u_i' > \bar{u}_i$, i transits to the state $(c^+, \bar{a}_i, \bar{u}_i)$.
- If $u_i' = \bar{u}_i$, i transits to the state $(c, \bar{a}_i, \bar{u}_i)$.

Hopeful: $z_i = (c^+, \bar{a}_i, \bar{u}_i)$. Agent i plays $\bar{a}_i$ and receives payoff u_i'.

- If $u_i' < \bar{u}_i$, i transits to the state $(c^-, \bar{a}_i, \bar{u}_i)$.
- If $u_i' > \bar{u}_i$, i transits to the state $(c, \bar{a}_i, \bar{u}_i')$.
- If $u_i' = \bar{u}_i$, i transits to the state $(c, \bar{a}_i, \bar{u}_i)$.

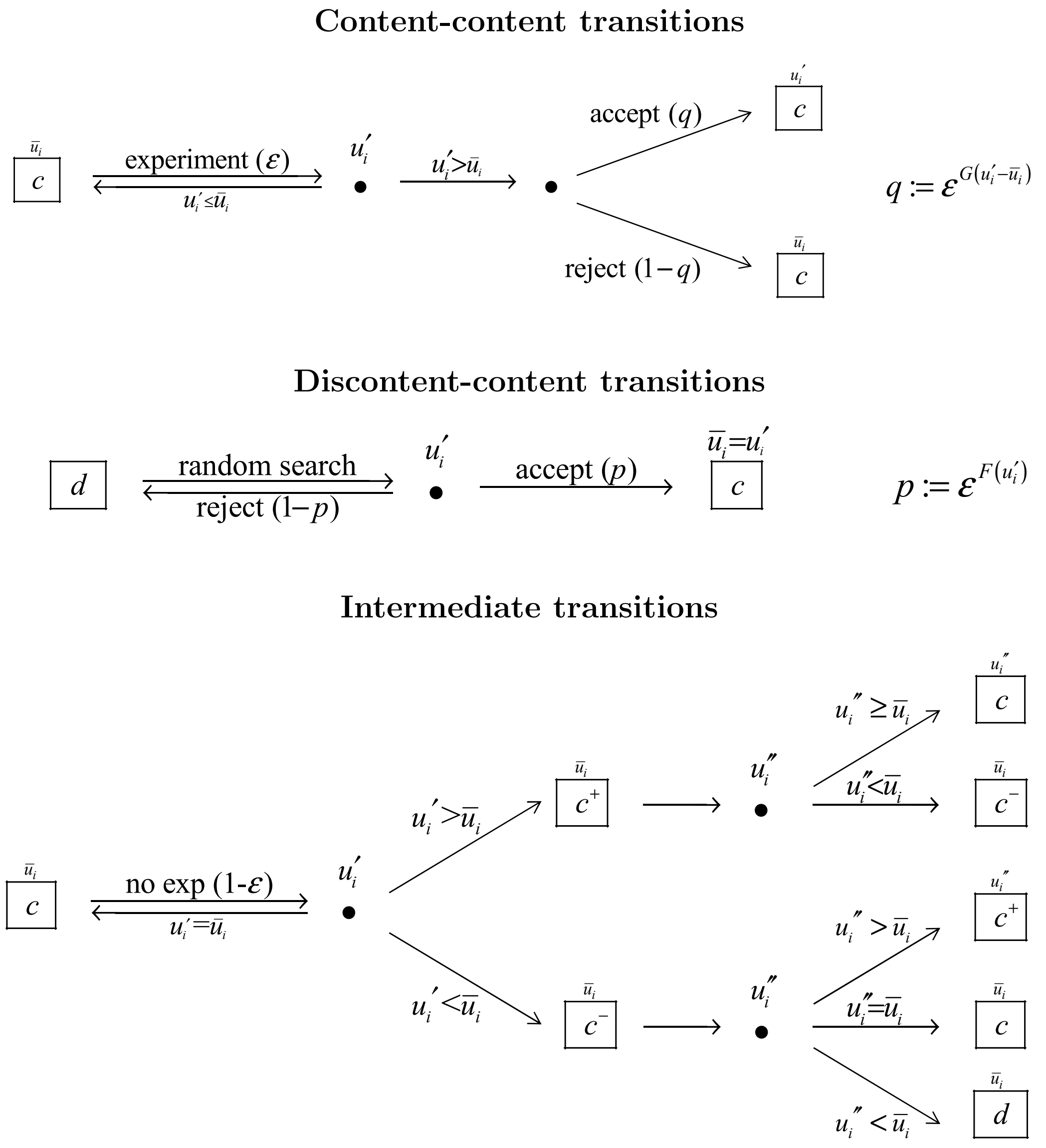
The logic of these intermediate states is that if i's realized payoff *confirms* his current mood, then i becomes discontent (from the watchful state) or content (from the hopeful state). If i's realized payoff does *not* confirm his current mood, i switches to the opposite intermediate state, or back to the content state if the realized payoff equals his benchmark payoff.

Discontent: $z_i = (d, \bar{a}_i, \bar{u}_i)$. Agent i plays an action a_i' drawn uniformly at random from A_i and receives payoff u_i'.

- i transits to the state $(c, \bar{a}_i', \bar{u}_i')$ with probability $\varepsilon^{F(u_i')}$; otherwise i remains in the state $(d, \bar{a}_i, \bar{u}_i)$.

Figure 2.1 illustrates the various transitions diagrammatically.

Figure 2.1: Transition diagram for active, matched agent (period $t+1$).



While the details of the learning algorithm are involved, the overall dynamics are qualitatively quite simple. The process oscillates between phases in which search is slow and search is fast.

1. When all players are content they occasionally experiment and search is slow. If the experiments lead to higher payoffs for some and no decrease for anyone, the players become content again with higher payoff benchmarks. This

is a hill-climbing phase in which payoffs are monotonically increasing.

2. A sequence of experiments by content players may result in someone's payoff going below his benchmark for two periods in a row. This player becomes discontent and starts searching every period. With some probability his searching causes other players to become discontent and they start searching. In effect, discontent search by one individual spreads through the population by contagion, creating high variance and leading to fast, wide-area search. This phase ends when the discontent players spontaneously settle down and become content again.

In section 2.5 we shall walk through an example step-by-step to show how the transition probabilities are computed in a concrete case. Before turning to these calculations, however, we shall discuss how this approach relates to earlier work and then state our main theorem.

2.3 Discussion

Learning rules that employ fast and slow search have been proposed in a variety of settings. In computer science, for example, there is a procedure known as WoLF (Win or Lose Fast) that has a qualitative flavor similar to the rule proposed above (Bowling and Veloso 2002). The basic idea is that when a player's payoff is high relative to realized average payoff he updates his strategy slowly, whereas he updates quickly if his payoff is low compared to the realized average. This approach is known to converge to Nash equilibrium in $2{\times}2$ games but not in general (Bowling and Veloso 2002). Similar ideas have been used to model animal foraging behavior (Houston et al. 1982, Motro and Shmida 1995, Thuijsman et al. 1995). For example, bees tend to search in the neighborhood of the last visited flower as long as the nectar yield is high, and search widely for an alternative patch otherwise (this is known as a near-far strategy). It would not be surprising if human subjects behaved in a similar fashion in complex learning environments, but to our knowledge this proposition has not been tested empirically.

In any event, we make no claim that the rule we have described is accurate from an empirical standpoint. The contribution of the present work is to demonstrate the existence of simple, completely uncoupled rules that select efficient equilibria in a wide class of games. This addresses an important problem in the application of game theory to distributed control, where the objective is to guarantee good system-wide performance using completely decentralized adaptive procedures. The issue of descriptive accuracy does not arise here, because one can build the learning rule into the agents by design. What matters is that the rule is simple to execute and requires little information. In our procedure it suffices to keep track of just three items – mood, benchmark action, and benchmark payoff – and to compare these with the received payoff each period.

2.4 Statement of the main result

To state our results in the most transparent form we shall impose certain restrictions on the acceptance functions F and G. In particular we shall assume that they are *strictly decreasing linear functions*

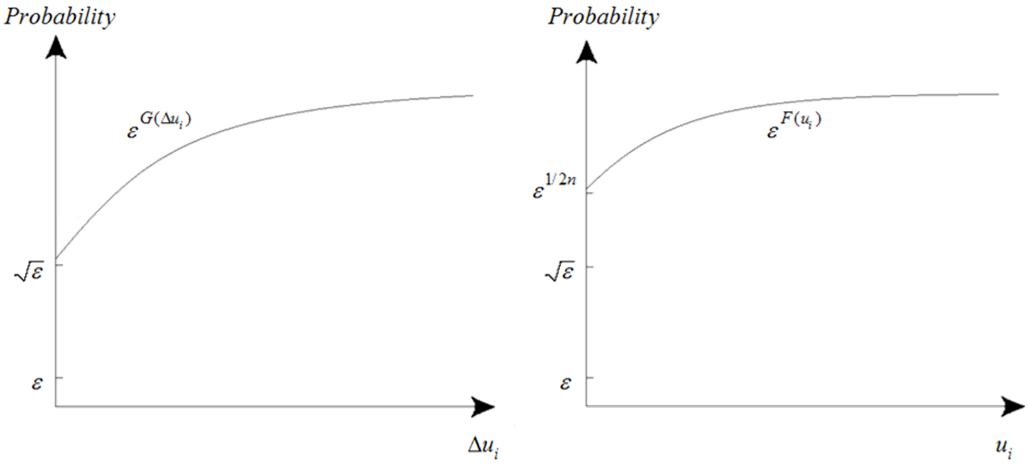
$$F(u) = -\varphi_1 \cdot u + \varphi_2, \text{ where } \varphi_1 > 0, \tag{2.1}$$

$$G(\Delta u) = -\gamma_1 \cdot \Delta u + \gamma_2, \text{ where } \gamma_1 > 0. \tag{2.2}$$

The coefficients φ_i, γ_i are chosen so that $F(u)$ and $G(\Delta u)$ are *strictly positive* for all $u \in U$ and $\Delta u \in \Delta U$, and are 'small' in the following sense:

$$0 < G(\Delta u) < \frac{1}{2} \text{ and } 0 < F(u) < \frac{1}{2n} \tag{2.3}$$

Figure 2.2: Acceptance probabilities for content agents (left panel) and discontent agents (tight panel).



This condition implies that the acceptance probabilities are fairly large relative to the probability of conducting an experiment. In particular the probability of accepting the outcome of an experiment ($\varepsilon^{G(\Delta u)}$) is always greater than $\sqrt{\varepsilon}$, and the probability of accepting the outcome of a random search is greater still (see Figure 2.2). We do not claim that these bounds are best possible, but they are easy to work with and yield sharp analytical predictions. The assumption of linearity will be useful for stating our first theorem, but in fact a more general equilibrium selection result holds even when the functions are nonlinear and differ among agents, as we shall show in section 2.7.

The learning rule described above is a variant of log linear learning that we shall call *log linear trial and error learning.* To see the connection, let $\varepsilon = e^{-\beta}$, where $\beta > 0$. Then the log probability of accepting the outcome of an experiment is

$$\log \mathbb{P}(\text{accept experiment with payoff gain } \Delta u > 0) = -\beta \log \varepsilon(\gamma_1 \Delta u - \beta \gamma_2). \tag{2.4}$$

Similarly, the log probability of accepting the outcome of a random search is

$$\log \mathbb{P}(\text{accept search with payoff } u) = -\beta \log \varepsilon(\varphi_1 u - \beta \varphi_2). \tag{2.5}$$

Notice that the probability of acceptance depends on the cardinality of the payoff gain or payoff level respectively. Hence the players' utilities cannot be rescaled or translated at will, as is the case with von Neumann-Morgenstern utility functions. The larger the payoff gain from an experiment, the higher the probability that it will be accepted. Similarly, the larger the payoff level that results from a random search, the higher the probability that the current action will be accepted.

Our equilibrium selection result will be framed in terms of two quantities: the welfare of a state and the stability of a state.

Welfare. The welfare of state $z = (m, \bar{a}, \bar{u})$ is the sum of the players' payoffs from their benchmark actions:

$$W(z) = \sum_{i=1}^{n} u_i(\bar{a}). \tag{2.6}$$

Stability. An action tuple $\bar{a} \in A$ is a δ-*equilibrium* for some $\delta \geq 0$ if

$$\forall i, \forall a_i \in A_i, \;\; u_i(a_i, \bar{a}_i) - u_i(\bar{a}) \leq \delta. \tag{2.7}$$

The stability of state $z = (m, \bar{a}, \bar{u})$ is the minimum $\delta \geq 0$ such that $\bar{a}$ is a

δ-equilibrium:

$$S(z) = \min\{\delta : \text{ the benchmark actions constitute a } \delta - \text{equilibrium}\}. \quad (2.8)$$

Note that the larger δ is, the less stable the state is, i.e., the greater is the incentive for some player to deviate. The following concepts will be needed to state our main result.

Stochastic stability. The set of *stochastically stable states* Z^* of the process $z(t)$ is the minimal subset of states such that, given any small $\alpha > 0$, there is a number $\varepsilon_\alpha > 0$ such that whenever $0 < \varepsilon < \varepsilon_\alpha$, $z(t) \in Z^*$ for at least $1 - \alpha$ of all periods t.

Interdependence. An n-person game $\mathcal{G}$ on the finite action space is *interdependent* if, for every $a \in A$ and every proper subset of players $\varnothing \subset J \subset N$, there exists some player $i \notin J$ and a choice of actions a'_J such that $u_i(a'_J, a_{N-J}) \neq u_i(a_J, a_{N-J})$.

In other words, given any current choice of actions $a \in A$, any proper subset of players J can cause a payoff change for some player not in J by a suitable change in their actions. Note that if a game has generic payoffs (and therefore no payoff ties), interdependence is automatically satisfied. However it is a much weaker condition. Consider, for example, a traffic game in which agents are free to choose any route they wish. There are many payoff ties because a local change of route by one player does not change the payoffs of players who are using completely different routes. But it satisfies the interdependence condition because a given player, or set of players, can (if they like) switch to a route that is being used by another player and thereby change his payoff.

Aligned. The benchmarks in state $z = (m, \bar{a}, \bar{u})$ are *aligned* if the benchmark payoffs result from playing the benchmark actions, that is, if $\bar{u}_i = u_i(\bar{a})$ for every player i.

Equilibrium state. A state z is an *equilibrium state* if all players are content, their benchmark actions constitute a Nash equilibrium, and their benchmark payoffs are aligned with their benchmark actions.

Theorem 3. *Let $\mathcal{G}$ be an interdependent n-person game on a finite action space A. Suppose that all players use log linear trial and error learning with experimentation probability ε and acceptance functions F and G satisfying the conditions in equations (2.1)-(2.3).*

1. *If $\mathcal{G}$ has at least one pure Nash equilibrium, then every stochastically stable state is an equilibrium state that maximizes $W(z)$ among all equilibrium states.*

2. *If $\mathcal{G}$ has no pure Nash equilibrium, every stochastically stable state maximizes $\varphi_1 W(z) - \gamma_1 S(z)$ among all $z \in C^0$.*

Notice that this result involves a cardinal comparison of the players' utilities. Players with large absolute utility are weighted heavily in the welfare function, and players with large utility differences are weighted heavily in the stability function. If a player's utility function were to be scaled up, he would count more heavily in both the welfare and the stability function. This would improve the likelihood of states in which this player has a high payoff, and decrease the likelihood of states in which this player has an incentive to deviate.

2.5 Examples

Before turning to the proof we shall consider two simple examples. The first illustrates the result when there is no pure Nash equilibrium. The second illustrates it when there are multiple Nash equilibria. In this case we shall walk through the algorithm to show why the Pareto-dominant equilibrium is selected instead of the risk-dominant equilibrium.

Example 1. Let $\mathcal{G}$ be a 2×2 game with payoff matrix

	A	B
A	$30, 30$	$0, 40$
B	$24, 23$	$10, 20$

This game has no pure Nash equilibria, so by Theorem 3 the learning process selects the combination that maximizes $\varphi_1 \cdot W - \gamma_1 \cdot S$. Figure 2.3 illustrates the case $\gamma_1/\varphi_1 = 1$ in which AA is selected.

In general, the outcome depends on the ratio γ_1/φ_1 as shown in Figure 2.4. Note that the welfare maximizing state AA is selected whenever γ_1/φ_1 is sufficiently small, that is, whenever the marginal change in the rate of acceptance by a discontent player is sufficiently large relative to the marginal change in the rate of acceptance of an experiment.

Figure 2.3: The tradeoff between welfare (W) and stability (S) in the 2×2 game of Example 1, which has no pure Nash equilibrium).

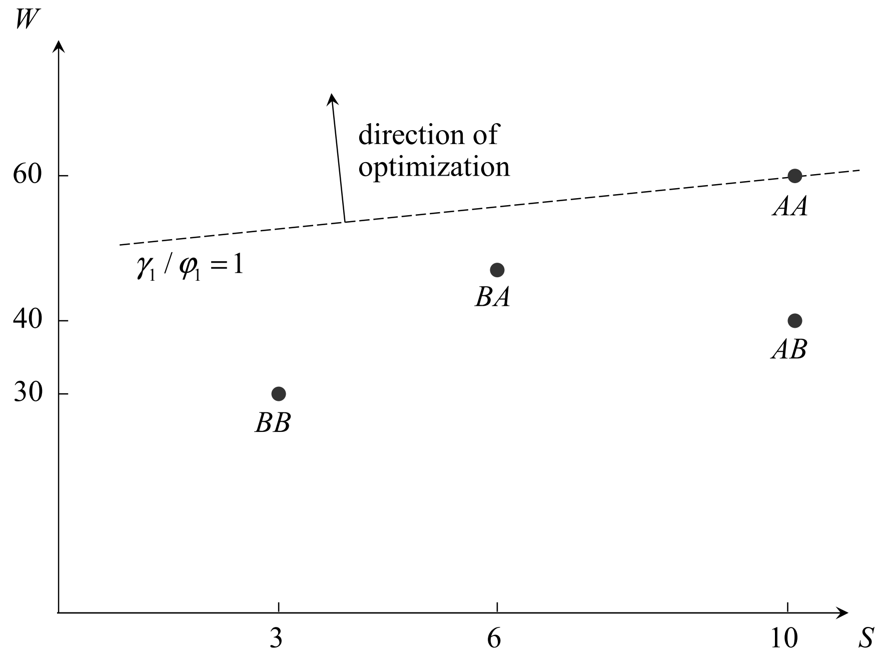
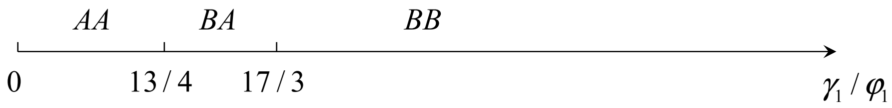


Figure 2.4: Stochastically stable outcomes for Example 1 as a function of γ_1/φ_1.



Example 2. Let $\mathcal{G}$ be a 2×2 coordination game with payoff matrix

	A	B
A	$40, 40$	$30, 10$
B	$10, 30$	$50, 50$

This game has two pure equilibria: AA and BB. The former is risk-dominant and the latter is Pareto-dominant. By Theorem 3 our learning process selects the Pareto-dominant equilibrium over the risk-dominant equilibrium. It therefore differs from many other learning processes in the evolutionary literature (Kandori et al. 1993, Young 1993a, Blume 1993, 1995, 2003). It also differs from the algorithm in Young (2009b), which provides no clear basis for discriminating between the two equilibria.

It is instructive to walk through the algorithm step-by-step to see how the selection mechanism works. For this purpose we need to consider specific acceptance functions. Let us choose the following:

$$F(u) = -0.001 \cdot u + 0.1, \quad G(\Delta u) = -0.01 \cdot \Delta u + 0.5$$

It is straightforward to verify that these functions satisfy the conditions in equations (2.1)-(2.3) for the game in question.

Let us begin by supposing that the learning process is currently in the state AB and that both players are content. In the next period it could happen that the column player experiments by playing action A (probability ε) and the row player does not experiment (probability $1 - \varepsilon$). In this case they play

AA, hence the column player experiences a payoff gain of 30 while the row player experiences a payoff gain of 10. Since the column player's experiment led to a payoff gain, he accepts his new action as his benchmark action with probability $\varepsilon^{G(\Delta u)} = \varepsilon^{G(30)} = \varepsilon^{0.2}$. Meanwhile the row player has turned hopeful because he experienced a non-self-triggered payoff gain. In the next period, the content column player continues to play his new benchmark action A with probability $1-\varepsilon$, and the hopeful row player plays his benchmark action A with probability 1. Therefore the row player experiences a higher payoff than his benchmark for two consecutive periods, which causes him to become content with 40 as his new benchmark payoff and A as his benchmark action. Overall this sequence of transitions from the all-content state AB to the all-content state AA has probability $\varepsilon \cdot (1-\varepsilon)^2 \cdot \varepsilon^{0.2}$ (see Figure 2.5).

Next let us consider the probability of exiting the all-content state BB. Note that any such transition involves a payoff decrease for both players. Hence for both of them to become content in another state, they must both accept lower benchmark payoffs. This can only happen if they first become discontent and subsequently become content again. Let D denote the set of states in which both players are discontent. We shall compute the probability of transiting from the all-content state BB to some state in D.

Since BB is a pure Nash equilibrium no unilateral deviation results in a payoff gain, and hence no single player will accept the outcome of an experiment. In fact, to move to D some player must experiment twice in a row. Suppose that next period the row player experiments and the column player does not and that this happens again in the following period. The probability of the first event is $\varepsilon \cdot (1-\varepsilon)$ while the probability of the second event is ε. (The reason for the latter is that, by the second period, the column player has become watchful and therefore does not experiment.) Since the column player experiences a payoff below his previous benchmark for two successive periods he becomes discontent. Now over the *next two periods* the probability is $0.5^2 = 0.25$ that the discontent column player chooses A both times, and that the row player does not experiment. This event has probability $0.25 \cdot (1-\varepsilon)$ because in the meantime the row player has become watchful. Overall this sequence of transitions from the all-content state BB to the set D has probability $0.25 \cdot \varepsilon^2 \cdot (1-\varepsilon)^2$.

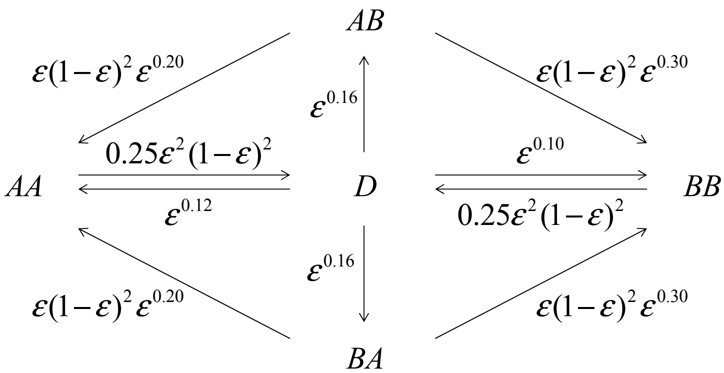
Next we turn to evaluating the probability of moving from a state in D to an all-content state. By assumption a discontent player spontaneously turns content with probability $\varepsilon^{F(u)}$ in any given period. (Note that this probability is *independent* of his benchmark payoff.) Therefore the probabilities of moving from D to each of the four all-content states are as follows:

$$D \xrightarrow{(\varepsilon^{F(40)})^2 = \varepsilon^{0.12}} AA, \quad D \xrightarrow{(\varepsilon^{F(30) \cdot F(10)}) = \varepsilon^{0.16}} AB,$$
$$D \xrightarrow{(\varepsilon^{F(10) \cdot F(30)}) = \varepsilon^{0.16}} BA, \quad D \xrightarrow{(\varepsilon^{F(50)})^2 = \varepsilon^{0.1}} BB$$

Note that among all possible ways of moving out of D, the move to all-content BB has highest probability.

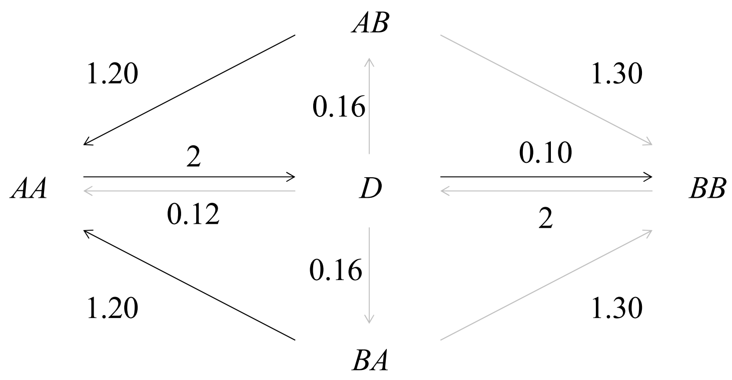
The situation is summarized in Figure 2.5, which shows the highest probability of transiting between D and each of the four all-content states (possibly via intermediate states, which are not shown).

Figure 2.5: Maximal transition probabilities between the four all-content states and the set D in which everyone is discontent.



The key to the analysis is the relative magnitude of the various transition probabilities when ε is very small. This is determined by the size of the exponent on ε, which is called the *resistance* of the transition. These resistances are shown in Figure 2.6.

Figure 2.6: Resistance of the transitions between states.



Note that the two equilibrium states AA and BB are the hardest to leave (resistance 2 in each case) and that equilibrium BB is easier to enter than AA. Thus on the face of it the process would seem to favor BB over AA, but this argument is insufficient. For example, it is easier to go from AB to AA (resistance 1.2) than from AB to BB (resistance 1.3), which suggests that BB might not be favored after all.

A basic fact about such processes is that it does not suffice to look at the 'local' resistance to entering and exiting different states; one must compare the resistances of various spanning trees in the graph. In the present example, the arrows shown in black in Figure 2.6 constitute one such tree: it has four edges and there is a unique directed path in the tree from every state other than BB to the state BB. (Such a tree is said to be rooted at BB.)

More generally, for each state z one must compute the minimum total resistance $\rho(z)$ over all spanning trees rooted at z, and then find the state such that $\rho(\cdot)$ is minimized The results are shown in Table 2.1 for each of the five states in the example. From this calculation we see that the minimum occurs uniquely at BB (with resistance 4.50), which implies that it is the unique stochastically stable state of the process. (This result is explained in greater generality in the next section.)

Table 2.1: Minimum resistance trees rooted at each state for the example of Figure 2.6.

$\rho(BB)$	=	$r(AA \to D, AB \to AA, BA\ toAA, D \to BB)$	=	4.50
$\rho(AA)$	=	$r(BB \to D, AB \to AA, BA\ toAA, D \to AA)$	=	4.52
$\rho(AB)$	=	$r(AA \to D, BB \to D, BA\ toAA, D \to AB)$	=	5.36
$\rho(BA)$	=	$r(AA \to D, BB \to D, AB\ toAA, D \to BA)$	=	5.36
$\rho(D)$	=	$r(AA \to D, BB \to D, AB\ toAA, BA \to AA)$	=	6.40

2.6 Proof of the main result

For given parameters ε, F, and G, the learning process defines a finite Markov chain on the state space Z. For every two states $z, z' \in Z$ there is a probability (possibly zero) of transiting from z to z' in one period. We shall write this one-period transition probability as a function of the experimentation rate ε: $p^{\varepsilon}_{zz'}$. Recall the definitions and results introduced in section 1.3 and the therein defined Program 2 to compute the set of stochastically stable states which we shall follow.

It is left to the reader to verify that the defined process satisfies the conditions for stochastic stability analysis to apply. The second step then is to characterize the recurrent classes of the unperturbed process P^0, that is, the process when $\varepsilon = 0$. In this situation, no one experiments and no one converts from being discontent to content.

Notation. Let Z^0 be the subset of states in which everyone's benchmarks are aligned. Let C^0 be the subset of Z^0 in which everyone is content, let E^0 be the subset of C^0 in which the benchmark actions constitute a pure Nash equilibrium and let E^* be the subset of E^0 on which welfare is maximized. Finally, let

$$D = \{\text{all states } z \in Z \text{ in which } \textit{every} \text{ player is discontent}\}. \tag{2.9}$$

Recall that if a player is discontent, he chooses an action according to a distribution that has full support and is independent of the current benchmark actions and payoffs. Moreover, the probability of accepting the outcome of such a search depends only on its realized payoff, and the old benchmarks are discarded. Hence, for any $w \in D$ and any $z \notin D$, the probability of the transition $w \to z$ is *independent* of w. Next consider a transition of form $z \to w$ where $w \in D$ and $z \notin D$. In this case the action and payoff benchmarks are the same in the two states, hence there is a unique $w \in D$ such that $z \to w$. We can therefore collapse the set D into a single state $\bar{D}$ and define the transition probabilities as follows:

$$\forall z \in Z \backslash D,\ \ \mathbb{P}(\bar{D} \to z) \ \equiv \ \mathbb{P}(w \to z) \quad \text{for all } w \in D \tag{2.10}$$

$$\forall z \in Z \backslash D,\ \ \mathbb{P}(z \to \bar{D}) \ \equiv \ \mathbb{P}(z \to w) \quad \text{for some } w \in D \tag{2.11}$$

Lemma 4. *The recurrent classes of the unperturbed process are $\bar{D}$ and all singletons $\{z\}$ such that $z \in C^0$.*

Proof. First we shall show that every element of C^0 is an absorbing state (a singleton recurrent class) of the unperturbed process ($\varepsilon = 0$). Suppose that $z \in C^0$, that is, the benchmarks are aligned and everyone is content. Since $\varepsilon = 0$, no one experiments and everyone replays his benchmark action next period with probability one. Hence the process remains in state z with probability one.

Next suppose that $z = \bar{D}$, that is, everyone is discontent. The probability that any given player becomes content next period is $\varepsilon^{F(\cdot)} = 0$. (Recall that $F(\cdot)$ is strictly positive on the domain of all payoffs that result from some feasible combination of actions.) Hence $\bar{D}$ is absorbing, i.e., a singleton recurrent class. (Recall that a discontent player chooses each of his available actions with a probability that is bounded away from zero and independent of ε.)

It remains to be shown that there are no other recurrent classes of the unperturbed process. We first establish the following.

Claim 5. *Given any state z in which at least one player is discontent, there exists a sequence of transitions in the unperturbed process to $\bar{D}$.*

Consider a state in which some player i is discontent. By interdependence he can choose an action that alters the payoff of someone else, say j. Assume that i plays this action for two periods in a row. If j's payoff decreases then in two periods he will become discontent also. If j's payoff increases then in two periods he will become content again with a higher payoff benchmark. At this point there is a positive probability that the first player, i, will revert to playing his original action for two periods. This causes player j's payoff to decrease relative to the new benchmark. Thus there is a positive probability that, in four periods or less, i's behavior will cause j to become discontent. (The argument implicitly assumed that no one except for i and j changed action in the interim, but this event also has positive probability.) It follows that there is a series of transitions to a state where both i and j are discontent. By interdependence there are actions of i and j that cause a third player to become discontent. The argument continues in this manner until the process reaches a state where all players are discontent, which establishes the claim.

An immediate consequence is that $\bar{D}$ *is the only recurrent state in which some player is discontent.* Thus, to conclude the proof of the lemma, it suffices to show that if z is a state in which no one is discontent, then there is a finite sequence of transitions to $\bar{D}$ or to C^0. Suppose that someone in z is in a transient mood (c^+ or c^-). Since no one is discontent and no one experiments, there is no change in the players' actions next period, and therefore no change in their realized payoffs. Hence everyone in a transient mood switches to c or d in one period. If anyone switches to d there is a series of transitions to $\bar{D}$. Otherwise everyone becomes content (or already was content) and their benchmarks are now aligned, hence the process has arrived at a state in C^0. ☺

We shall now move to the third point of Program 2 in section 1.3, namely of computing the resistances between recurrent classes.

As introduced in section 1.3 we know from Young (1993a, Theorem 4) that the computation of the stochastically stable states can be reduced to an analysis of rooted trees on the vertex set R consisting solely of the recurrent classes. By Lemma 4, R consists of the singleton states in C^0, and also the singleton state $\{\bar{D}\}$. Henceforth we shall omit the set brackets $\{\}$ to avoid cumbersome notation.

r^*-function. Define the function $r^*(z)$ as follows

$$\forall z \in R, \quad r^*(z) = \min\{r(z, w) : w \in R - \{z\}\}. \tag{2.12}$$

Easy. An *easy* edge from a state z is an edge $z \to w$, $w \neq z$, such that $r(z \to w) = r^*(z)$. An *easy path* is a sequence $w^1 \to w^2 \to \ldots \to w^m$ in which each edge is easy and all states are distinct. An *easy tree* is a tree all of whose edges are easy.

This idea was illustrated in Figure 2.6, where the easy edges are shown in black and the other edges in gray. Together the four easy edges form an easy tree that is rooted at the state BB. This identifies BB as the stochastically stable state in this particular case. While it is not always possible to determine the stochastically stable state in this way, the identification of the easy edges is a key first step in the analysis. For this purpose we need to evaluate the function r^* on the various recurrent classes. This is the purpose of the next three lemmas.

Lemma 6. *$\forall e \in E^0$, $r^*(e) = 2$ and $e \to \bar{D}$ is an easy edge.*

Proof. Let $e = (m\bar{a}, \bar{u}) \in E^0$, where $\bar{a}$ is a pure Nash equilibrium. Consider any outgoing edge $e \to z$ where $z \in C^0$ and $z \neq e$. Since e and z are distinct, everyone is content, and the benchmark payoffs and actions are aligned in both states, they must differ in their benchmark actions. Now consider any path from e to z in the full state space Z. Along any such path at least two players must experiment with new actions (an event with probability $O(\varepsilon^2)$), or one player must experiment twice in succession (also an event with probability $O(\varepsilon^2)$) in order for someone's benchmark action or payoff to change. (A single experiment is not accepted by the experimenter because $\bar{a}$ is a Nash equilibrium, so the payoff from an experiment does not lead to a payoff gain. Furthermore, although some other players may temporarily become hopeful or watchful, they revert to being content with their old benchmarks unless a second experiment occurs in the interim.) It follows that $r(e \to z) \geq 2$.

It remains to be shown that the resistance of the transition $e \to \bar{D}$ is exactly two, which we shall do by constructing a particular path from e to $\bar{D}$ in the full state space. Choose some player i. By interdependence there exists an action $a_i \neq \bar{a}_i$ and a player $j \neq i$ such that i's change of action affects j's payoff: $u_j(a_i, \bar{a}_{-i}) \neq u_j(\bar{a})$. Let player i experiment by playing a_i twice in succession, and suppose that no one else experiments at the same time. This event has probability $O(\varepsilon^2)$, so the associated resistance is two. If $u_j(a_i, \bar{a}_{-i}) > u_j(\bar{a})$, player j's mood changes to c^+ after the first experiment and to c again after the second experiment. Note that at this point j has a new higher benchmark, namely, $u_j(a_i, \bar{a}_{-i})$. Now with probability $(1-\varepsilon)^{2n}$ player i reverts to playing $\bar{a}_i$ for the next two periods and no one else experiments during these periods. This causes j to become discontent. By Claim 5, there is a zero-resistance path to the all-discontent state. The other case, namely $u_j(a_i, \bar{a}_{-i}) < u_j(\bar{a})$, also leads to an all-discontent state with no further resistance. We have therefore shown that $r(e \to \bar{D}) = 2$, and hence that $e \to \bar{D}$ is an easy edge. ☺

Lemma 7. *$\forall z \in C^0 - E^0$, $r^*(z) = 1 + G(S(z))$, and if $z \to z'$ is an easy edge with $z' \in C^0$, then $W(z) < W(z')$.*

Proof. Let $z = (m\bar{a}, \bar{u}) \in C^0 \backslash E^0$, in which case $\bar{a}$ is not a Nash equilibrium. Then there exists an agent i and an action $a_i \neq \bar{a}_i$ such that $u_i(a_i, a_{-i} > u_i(\bar{a}) = \bar{u}_i$. Among all such agents i and actions a_i suppose that $\Delta u_i = u_i(a_i, \bar{a}_{-i}) - u_i(\bar{a})$ is a maximum. Let i experiment once with this action and accept the outcome of the experiment, and suppose that no one else experiments at the same time. The probability of this event is $O(\varepsilon^{1+G(\Delta u_i)})$. If the experiment causes everyone else's payoff to stay the same or go up, and if no one experiments in the next period (with the latter event having probability $(1-\varepsilon)^n$), then a state z' is reached after one period in which everyone is content, the benchmarks are aligned ($z' \in C^0$) and $W(z) < W(z')$. The total resistance of this path is $r(z \to z') = 1 + G(S(z))$. (Recall from equation (2.8) that $S(z)$ is the largest $\delta > 0$ such that someone can gain δ by a unilateral deviation.)

The other possibility is that the experiment causes someone else's payoff to *decrease*, in which case there is a zero-resistance series of transitions to the state $\bar{D}$. Hence in this case we have $r(z \to \bar{D}) = 1 + G(S(z))$. We conclude that, if there is only one experiment, the process either transits to a recurrent class $z' \in C^0$ satisfying $W(z') > W(z)$, or it transits to $\bar{D}$. In both cases the resistance of the transition is $1 + G(S(z))$.

If there are two or more experiments, the resistance is at least two. By assumption, however, $G(\cdot) < 1/2$ (equation (2.3)), hence making two experiments has a higher resistance than making one experiment and accepting the outcome (the latter has resistance $1 + G(S(z)) < 1.5$). It follows that $r^*(z) = 1 + G(S(z))$, and if $z \to z'$ is an easy edge with $z' \neq \bar{D}$, then $W(z) < W(z')$. ☺

Lemma 8. *$\forall z \in C^0$, $r(\bar{D} \to z) = \sum_i F(u_i(\bar{a}))$, and $r^*(\bar{D}) = \min_{a \in A} \sum_i F(u_i(a))$.*

Proof. Let $\bar{D} \to z = (m\bar{a}, \bar{u}) \in C^0$. Recall that the transition probability $\mathbb{P}(w \to z)$ is the same for all $w \in D$, hence take any state $w^1 \in D$ as the starting point. The probability is $O(1)$ that next period every player i chooses $\bar{a}_i$, in which case their realized payoffs are $\bar{u}_i = u_i(\bar{a})$. They all accept these actions and payoffs as their new benchmarks with probability $\prod_i \varepsilon^{F(u_i(\bar{a}))} = \varepsilon^{\sum_i F(u_i(\bar{a}))}$, hence $r(\bar{D} \to z) \leq \sum_i F(\bar{u}_i)$.

We claim that in fact $r(\bar{D} \to z) = \sum_i F(\bar{u}_i)$. Let $z = (m, \bar{a}, \bar{u}) \in C^0$ and consider a least-resistant path $w^1, w^2, \ldots, w^m = z$. For each player i there must be some time in the sequence where i was discontent and accepted a benchmark payoff that was $\bar{u}_i$ *or less.* (There may also have been a time when i was discontent and accepted a benchmark payoff that was strictly more than $\bar{u}_i$, but in that case there must have been a later time at which he accepted a payoff that was $\bar{u}_i$ or less, which means he must have been discontent.) The probability of such an acceptance is at most $\varepsilon^{F(\bar{u}_i)}$, because the probability of acceptance is increasing in u_i. This reasoning applies to every player, hence the total resistance of this path from $\bar{D}$ to z must be at least $\sum_i F(\bar{u}_i)$. This proves that $r(\bar{D} \to z) = \sum_i F(\bar{u}_i)$. The claim that $r^*(\bar{D}) = \min_{a \in A} \sum_i F(u_i(a))$ follows from the fact that F is monotone decreasing. ☺

We shall now move to the fourth point of Program 2 in section 1.3, namely of computing the stochastic potential of each recurrent class.

Lemma 9. *There exists a $\bar{D}$-tree T^*_D that is easy.*

Proof. We shall conduct the proof on transitions between recurrent classes, each of which forms a node of the graph. Choose a node $z \neq \bar{D}$ and consider an easy outgoing edge $z \to \cdot$. If there are several such edges choose one that points to $\bar{D}$, that is, choose $z \to \bar{D}$ if it is easy. This implies in particular that for every $e \in E^0$ we select the edge $e \to \bar{D}$. (This follows from Lemma 6.)

We claim that the collection of all such edges forms a $\bar{D}$-tree. To establish this it suffices to show that there are no cycles. Suppose by way of contradiction that $z^1 \to z^2 \to \ldots \to z^m \to z^1$ is a shortest cycle. This cycle cannot involve any node in E^0, because by construction the outgoing edge from any such edge points towards $\bar{D}$, which has no outgoing edge. Therefore all $z^k \in C^0 - E^0$. Since all of the edges $z^k \to z^{k+1}$ are easy, Lemma 7 implies that $W(z^k) < W(z^{k+1})$. From this we conclude $W(z^1) < W(z^m) < W(z^1)$, which is impossible. ☺

Let

$$\rho^* = \rho(\bar{D}) = r(T^*_D) \tag{2.13}$$

Lemma 10. *For every $z \in C^0$ let $z \to w_z$ be the unique outgoing edge from z in T^*_D and define*

$$T^*_z = T^*_D \textit{ with } z \to w_z \textit{ removed and } \bar{D} \to z \textit{ added.} \tag{2.14}$$

T^*_z is a z-tree of least resistance and

$$\rho(z) = \rho^* - r(z \to w_z) + r(\bar{D} \to z). \tag{2.15}$$

Proof. Plainly, the tree T_z^* defined in equation (2.14) is a z-tree. Furthermore all of its edges are easy except possibly for the edge $\bar{D} \to z$. Hence it is a least-resistant z-tree among all z-trees that contain the edge $\bar{D} \to z$. We shall show that in fact it minimizes resistance among all z-trees.

Let T_z be some z-tree with minimum resistance, and suppose that it does not contain the edge $\bar{D} \to z$. Since it is a spanning tree it must contain some outgoing edge from $\bar{D}$, say $\bar{D} \to z'$. We can assume that $r(\bar{D} \to z') < r(\bar{D} \to z)$, for otherwise we could simply take out the edge $\bar{D} \to z'$ and put in the edge $\bar{D} \to z$ to obtain the desired result.

Let $u_1, u_2, \ldots, u_n$ be the benchmark payoffs in z and let be the benchmark payoffs in . By Lemma 8,

$$r(\bar{D} \to z) = \sum_i F(u_i) \text{ and } r(\bar{D} \to z') = \sum_i F(u_i'). \tag{2.16}$$

Since $r(\bar{D} \to z') < r(\bar{D} \to z)$ and $F(u_i)$ is by assumption decreasing in u_i, there must be some i such that $u_i' > u_i$. Let $I = \{i : u_i' > u_i\}$. Consider the unique path in T_z that goes from z' to z, say $z' = z^1, z^2, \ldots, z^m = z$, where each z^k is in R and they are distinct. Each edge $z^k \to z^{k+1}$ corresponds to a sequence of transitions in the full state space Z, and the union of all these transitions constitutes a path in Z from z' to z. Along this path, each player $i \in I$ must eventually lower his payoff benchmark because his starting benchmark u_i is greater than his ending benchmark u_i'. Thus at some point i must become discontent and adopt a new benchmark that is u_i or lower. Assume that this happens in the transition from z^{k_i} to z^{k_i+1}. Since $F(u_i)$ is strictly decreasing, the probability of adopting a benchmark that is u_i or lower is *at most* $\varepsilon^{F(u_i)}$. Hence, along the path from z^{k_i} to z^{k_i+1} someone experiments and accepts, and at some stage player i becomes discontent and accepts a payoff that is u_i or lower. The first event has resistance at least $1 + G(S(z^{k_i}))$ and the second has resistance at least $F(u_i)$. Therefore

$$r(z^{k_i} \to z^{k_i+1}) \geq 1 + G(S(z^{k_i})) + F(u_i). \tag{2.17}$$

We know from Lemma 7 that the least resistance of a path out of z^{k_i} is $r^*(z^{k_i}) = 1 + G(S(z^{k_i}))$. Hence $r(z^{k_i} \to z^{k_i+1}) \geq F(u_i) + r^*(z^{k_i})$. It follows that the total resistance along the sequence $z' = z^1, z^2, \ldots, z^m = z$ satisfies

$$\sum_{k=1}^{m-1} r(z^k \to z^{k+1}) \geq \sum_{i \in I} F(u_i) + \sum_{k=1}^{m-1} r^*(z^k). \tag{2.18}$$

The resistance of the edge $\bar{D} \to z'$ satisfies

$$r(\bar{D} \to z') = \sum_i F(u_i') > \sum_{i \in I} F(u_i') \geq \sum_{i \in I} F(u_i). \tag{2.19}$$

Hence in the tree T_z the outgoing edges from the nodes $\{\bar{D}, z^1, z^2, \ldots, z^{m-1}\}$ have a total resistance that is strictly greater than $\sum_i F(u_i) + \sum_{k=1}^{m-1} r^*(z^k)$. But in the tree T_z^* the edges from these nodes have a total resistance equal to $\sum_i F(u_i) + \sum_{k=1}^{m-1} r^*(z^k)$. Furthermore, at every other node the resistance of the outgoing edge in T_z is at least as great as it is in T_z^* (because the latter consists of easy edges). We conclude that T_z^* must minimize resistance among all z-trees, which completes the proof of the lemma. ☺

To complete the proof of Theorem 3, we shall first prove that $\forall e^* \in E^*$, $\forall e \in E^0 \backslash E^0$, $\forall z \in C^0 \backslash E^0$, we have

$$\rho(e^*) \overset{(i)}{<} \rho(e) \overset{(ii)}{<} \rho(z) \overset{(iii)}{<} \rho(\bar{D}) = \rho^*. \tag{2.20}$$

(Recall that E^* consists of those equilibrium states $e \in E^0$ that maximize $W(e)$.) Let $e = (m, \bar{a}, \bar{u}) \in E^0$. By construction $e \to \bar{D}$ is an edge in the easy tree $T_{\bar{D}}^*$ (see the beginning of the proof of Lemma 9), so equation (2.15) implies that

$$\rho(e) = \rho^* - r(e \to \bar{D}) + r(\bar{D} \to e). \tag{2.21}$$

From Lemma 6 we know that $r(e \to \bar{D}) = 2$. From Lemma 8 we know that

$$r(\bar{D} \to e) = \sum_i F(u_i(\bar{a})) = -\varphi_1 W(e) + n\varphi_2. \tag{2.22}$$

From this and equation (2.21) we conclude that

$$\rho(e) = \rho^* - 2 - \varphi_1 W(e) + n\varphi_2. \tag{2.23}$$

Now suppose that $e^* \in E^*$ and $e \in E^0 \backslash E^*$. Then $W(e^*) < W(e)$, so from equation (2.23) we conclude that $\rho(e^*) < \rho(e)$. This establishes (i).

To prove (ii), let $e \in E^0 \backslash E^*$ and $z \in C^0 \backslash E^0$. Recall from equation(2.15) that $\rho(z) = \rho^* - r(z \to w_z) + r(\bar{D} \to z)$, where $z \to w_z$ is an easy edge. Since $z \in C^0 - E^0$, we know from Lemma 7 that $r(z \to w_z) = 1 + G(S(z))$ and from Lemma 8 that $r(\bar{D} \to z) = -\varphi_1 W(z) + n\varphi_2$. Hence

$$\forall z \in C^0 \backslash E^0,\ \rho(z) = \rho^* - 1 - G(S(z)) - \varphi_1 W(z) + n\varphi_2. \tag{2.24}$$

Since $e \in E^0$, equation (2.23) implies that

$$\rho(e) = \rho^* - 2 - \varphi_1 W(e) + n\varphi_2. \tag{2.25}$$

Comparing equation (2.24) and (2.25) we see that $\rho(e) < \rho(z)$ provided

$$\varphi_1[W(z) - W(e)] < 1 - G(S(z)). \tag{2.26}$$

The right-hand side of equation (2.26) is greater than 1/2 because we assumed that $G(S(z)) < 1/2$ for all z (see equation (2.3)). The left-hand side is the sum of n differences $\sum_i F(u_i) - F(u_i')$, which is smaller than 1/2 because we assumed that $0 < F(\cdot) < 1/(2n)$ (equation (2.3) again). Hence equation (2.26) holds and therefore $\rho(e) < \rho(z)$, which proves (ii).

To prove (iii), observe that equation (2.3) implies

$$-\varphi_1 W(z) + n\varphi_2 \leq n \cdot max_u F(u) < 1/2. \tag{2.27}$$

From this and equation (2.24 it follows that $\rho(z) < \rho^*$. This completes the proof of equation (2.20).

We shall now turn to the two cases of the theorem: a pure Nash equilibrium exists and a pure Nash equilibrium does not exist.

Case 1. A pure Nash equilibrium exists ($E^0 \neq \varnothing$). By equation (2.20) we know that the welfare maximizing equilibrium states $e^* \in E^*$ minimize stochastic potential among all recurrent classes, hence by Theorem 1 they constitute the stochastically stable states. This establishes statement (i) of Theorem 3.

Case 2. No pure equilibrium exists ($E^0 = \varnothing$).

In this case the chain of inequalities (2.20) reduces to (iii), because there are no equilibrium states. In particular, $\bar{D}$ *cannot* be stochastically stable. It follows from equation (2.24) that the stochastically stable states are precisely those $z \in C^0$ that minimize

$$\begin{aligned} \rho(z) &= \rho^* - 1 - G(S(z)) - \varphi_1 W(z) + n\varphi_2 && (2.28) \\ &= \gamma_1 S(z) - \varphi_1 W(z) + (n\varphi_2 + \rho^* - 1) && (2.29) \end{aligned}$$

which is equivalent to maximizing $\varphi_1 W(z) - \gamma_1 S(z)$. This concludes the proof of Theorem 3.

2.7 Heterogeneous agents

Thus far we have assumed that ε, F and G are common to all agents and that F and G are linear. In this section we generalize Theorem 3 to the situation where the agents have heterogeneous learning parameters and the acceptance functions may be nonlinear. First let us consider what happens when agents have different rates of experimentation. Suppose that agent i has experimentation rate $\varepsilon_i = \lambda_i \varepsilon$ where λ_i is idiosyncratic to i and $\varepsilon > 0$ is common across agents. Varying ε changes the overall amount of noise in the search process while holding the ratio of the search rates fixed. Assume that agent i accepts the outcome of a random search with probability $\varepsilon_i^{F(u_i)} = (\lambda_i \varepsilon)^{F(u_i)}$, and accepts the outcome of an experiment with probability $\varepsilon_i^{G(\Delta u_i)} = (\lambda_i \varepsilon)^{G(\Delta u_i)}$. The introduction of the λ_i's does not alter the resistances, which are determined by the exponents of ε in the transition probabilities. Hence the stochastically stable states remain exactly as before, and Theorem 3 holds as stated.

Now suppose that i's probability of accepting the outcome of a random search is governed by some function $F_i(u_i)$, and i's probability of accepting the outcome of an experiment is governed by $G_i(\Delta u_i)$. Assume that these acceptance functions are strictly monotone decreasing but not necessarily linear, and that they satisfy the bounds in equation (2.3), that is,

$$0 < G_i(\Delta u_i) < 1/2 \text{ and } 0 < F_i(u_i) < 1/(2n) \text{ for all } i. \tag{2.30}$$

Given a state $z = (m, \bar{a}, \bar{u})$ let us redefine the welfare function as follows:

$$\widetilde{W}(z) = -\sum_i F_i(u_i(\bar{a})) \tag{2.31}$$

Recall that an equilibrium is *efficient* (Pareto-undominated) if there is no other equilibrium in which someone is better off and no one is worse off. Since the functions F_i are monotone decreasing, $\widetilde{W}(z)$ is monotone increasing in each player's utility $u_i(\bar{a})$. It follows that, among the equilibrium states, $\widetilde{W}(\cdot)$ is maximized at an efficient equilibrium. (If the F_i are linear, $\widetilde{W}(\cdot)$ maximizes a weighted sum of the agents' utilities.)

Next let us define

$$S_i(z) = S_i(m, \bar{a}, \bar{u}) = \max_{a_i \in A_i} \{u_i(a_i, \bar{a}_{-i}) - u_i(\bar{a}) : u_i(a_i, \bar{a}_{-i}) - u_i(\bar{a}) > 0\} \quad (2.32)$$

Further let $\widetilde{S}_i(z) = -G_i(S_i(z))$ and define the stability of a state z to be

$$\widetilde{S}(z) = \max_i \left\{\widetilde{S}_i(z)\right\}. \quad (2.33)$$

The larger $\widetilde{S}(z)$ is, the more likely it is that some player will accept the outcome of an experiment, hence the more prone the state is to being destabilized. A straightforward modification in the proof of Theorem 3 leads to the following:

Theorem 11. *Let $\mathcal{G}$ be an interdependent n-person game on a finite joint action space A. Suppose that each player i uses a learning rule with experimentation probability $\varepsilon_i = \lambda_i \varepsilon$ and monotone decreasing acceptance functions F_i and G_i satisfying the bounds in equation (2.30).*

1. *If the game has a pure Nash equilibrium then every stochastically stable state is an equilibrium state that maximizes $\widetilde{W}(z)$ among all equilibrium states, and hence is efficient.*

2. *If the game has no pure Nash equilibrium, the stochastically stable states maximize $\widetilde{W}(z) - \widetilde{S}(z)$ among all $z \in C^0$.*

2.8 Conclusion

In this chapter we have identified a completely uncoupled learning rule that selects an efficient pure equilibrium in any n-person game with generic payoffs that possesses at least one pure equilibrium. This provides a solution to an important problem in the application of game theory to distributed control, where the object is to design a system of autonomous interacting agents that optimizes some criterion of system-wide performance using simple feedback rules that require no information about the overall state of the system. The preceding analysis shows that this can be accomplished by a variant of log linear learning and two different search modes – fast and slow. Theorem 3 shows that by choosing the probabilities of experimentation and acceptance within an appropriate range, the process is in an efficient equilibrium a high proportion of the time. This allows one to reduce the price of anarchy substantially, because one need only compare the maximum welfare state to the maximum welfare equilibrium state. As an extra dividend we obtain a simple criterion for the selection of disequilibrium states. This criterion weights total welfare positively and the incentive to deviate negatively. As we have shown by example this concept differs from risk dominance.

It is, of course, legitimate to ask how long it takes for the learning rule to reach an efficient equilibrium from arbitrary initial conditions. We know in general that it can take an exponentially long time for an uncoupled learning process to converge to Nash equilibrium for games of arbitrary size (Hart and Mansour 2010); it can also take an exponentially long time to converge to a Pareto efficient, individually rational state (Babichenko 2014). The method that we used to prove Theorem 3 involved taking the experimentation rate to zero, in which case it takes a very long time in expectation for the process to reach the stochastically stable state(s). However, this does not rule out the possibility that the learning process *strongly favors* the stochastically stable states even when the experimentation probability is large; indeed this has been shown for 2×2 games played by large populations of players who make mistakes independently (Young 1998, Chapter 4). Thus it is possible that the learning algorithm with *intermediate* levels of experimentation would lead quite quickly to an approximate equilibrium, at least for some classes of games. This is a challenging open problem that we shall not attempt to tackle here.

Finally, we should note that the type of learning rules we have proposed are not meant to be empirically descriptive of how humans actually behave in large decentralized environments. Our aim has been to show that it is theoretically possible to achieve efficient equilibrium selection using simple, completely uncoupled rules. Nevertheless, it is conceivable that certain qualitative features of these rules are reflected in actual behavior. At the micro level, for example, one could test whether agents engage in different types of search (fast and slow) depending on their recent payoff history. One could also estimate the probability that they accept the outcome of a search as a function of their realized payoffs. At the macro level, one could examine whether agents playing a congestion game converge to a local maximum of the potential function, to a Pareto-optimal equilibrium, or fail to come close to equilibrium in any reasonable amount of time. Whether or not our learning model proves to have features that are descriptively accurate for human agents, the approach does suggest some testable questions that to the best of our knowledge have not been examined before.

Chapter 3

Evolutionary dynamics and equitable core selection in assignment games

3.1 Introduction

Many matching markets are decentralized and agents interact repeatedly with very little knowledge about the market as a whole. Examples include online markets for bringing together buyers and sellers of goods, matching workers with firms, matching hotels with clients, and matching men and women. In such markets matchings are repeatedly broken, reshuffled, and restored. Even after many encounters, however, agents may still have little information about the preferences of others, and they must experiment extensively before the market stabilizes.

In this chapter we propose a simple adaptive process that reflects the participants' limited information about the market. Agents have aspiration levels that they adjust from time to time based on their experienced payoffs. Matched agents occasionally experiment with higher bids in the hope of extracting more from another match, while single agents occasionally lower their bids in the hope of attracting a partner. There is no presumption that market participants or a central authority know anything about the distribution of others' preferences or that they can deduce such information from prior rounds of play. Instead they follow a process of trial and error in which they adjust their bids and offers in the hope of increasing their payoffs. Such aspiration adjustment rules are rooted in the psychology and learning literature.[1] A key feature of the rule we propose is that an agent's behavior does not require any information about other agents' actions or payoffs: the rule is completely uncoupled. It is therefore particularly well-suited to environments such as decentralized online markets where players interact anonymously and trades take place at many different prices. We shall show that this simple adaptive process leads to equitable solutions inside the core of the associated *assignment game* (Shapley and Shubik 1972). In particular, core stability and equity are achieved even though agents have no knowledge of the other agents' strategies or preferences, and there is no ex ante preference for equity.

The chapter is structured as follows. The next section discusses the related literature on matching and core implementation. Section 3.3 formally introduces assignment games and the relevant solution concepts. Section 3.4 describes the process of adjustment and search by individual agents. In sections 3.5 and 3.6 we show that the stochastically stable states of the process lie inside the core. Section 3.7 concludes with several open problems.

[1] There is an extensive literature in psychology and experimental game theory on trial and error and aspiration adjustment. See in particular the learning models of Thorndike (1898), Hoppe (1931), Estes (1950), Bush and Mosteller (1955), Herrnstein (1961), and aspiration adjustment and directional learning dynamics of Heckhausen (1955), Sauermann and Selten (1962), Selten and Stoecker (1986), Selten (1998).

3.2 Related literature

Our results fit into a growing literature showing how cooperative game solutions can be implemented via noncooperative dynamic learning processes (Agastya 1997, 1999, Arnold and Schwalbe 2002, Newton 2010, 2012, Rozen 2013, Sawa 2014). A particularly interesting class of cooperative games are assignment games, in which every potential matched pair has a cooperative 'value'. Shapley and Shubik (1972) showed that the core of such a game is always nonempty.[2] Subsequently various authors have explored refinements of the assignment game core, including the kernel (Rochford 1984) and the nucleolus (Huberman 1980, Solymosi and Raghavan 1994, Nunez 2004, Llerena et al. 2015). To the best of our knowledge, however, there has been no prior work showing how a core refinement is selected via a decentralized learning process, which is the subject of the present chapter.

This chapter establishes convergence to the core of the assignment game for a class of natural dynamics and selection of a core refinement under payoff perturbations. We are not aware of prior work comparable with our selection result. There are, however, several recent papers that also address the issue of core convergence for a variety of related processes (Chen et al. 2011, Biró et al. 2012, Klaus and Payot 2013, Bayati et al. 2014). These processes are different from ours, in particular they are not aspiration-adjustment learning processes, and they do not provide a selection mechanism for a core refinement as we do here. The closest relative to our work is the concurrent paper by Chen et al. (2011), which demonstrates a decentralized process where, similarly as in our process, pairs of players from the two market sides randomly meet in search of higher payoffs. This process also leads almost surely to solutions in the core.

[2]Important subsequent papers include Crawford and Knoer (1981), Kelso and Crawford (1982), Demange and Gale (1985), and Demange et al. (1986).

Chen et al. (2011) and our work are independent and parallel work. They provide a constructive proof based on their process which is similar to ours for the proof of the convergence theorem. Thus, theirs as well as our algorithm (proof of Theorem 12) can be used to find core outcomes. Biró et al. (2012) generalizes Chen et al. (2011) to transferable-utility roommate problems. In contrast to Chen et al. (2011) and our proof, Biró et al. (2012) use a target argument which cannot be implemented to obtain a core outcome. Biró et al. (2012)'s proof technique is subsequently used in Klaus and Payot (2013) to prove the result of Chen et al. (2011) for continuous payoff space in the assignment game. A particularity in this case is the fact that the assignment may continue to change as payoffs approximate a core outcome. Finally, Bayati et al. (2014) study the rate of convergence of a related bargaining process for the roommate problem in which players know their best alternatives at each stage. The main difference of this process to ours is that agents best reply (i.e. they have a lot of information about their best alternatives). Moreover, the order of activation is fixed, not random, and matches are only formed once a stable outcome is found.[3] An important feature of our learning process is that it is explicitly formulated in terms of random bids of workers and random offers of firms (as in the original set-up by Shapley and Shubik 1972), which allows a completely uncoupled set-up of the dynamic.

There is also a related literature on the marriage problem (Gale and Shapley 1962). In this setting the players have ordinal preferences for being matched with members of the other population, and the core consists of matchings such that no pair would prefer each other to their current partners.[4] Typically, many matchings turn out to be stable. Roth and Vande Vate (1990) demonstrate a random blocking pair dynamic that leads almost surely to the core in such

[3]In chapter 4 we discuss the differences to our set-up in more detail. We then investigate the convergence rate properties of a process closely related to the current model.

[4]See Roth and Sotomayor (1992) for a text on two-sided matching.

games. Chung (2000), Diamantoudi et al. (2004) and Inarra et al. (2008, 2013) establish similar results for nontransferable-utility roommate problems, while Klaus and Klijn (2007) and Kojima and Ünver (2008) treat the case of many-to-one and many-to-many nontransferable-utility matchings. Another branch of the literature considers stochastic updating procedures that place high probability on core solutions, that is, the stochastically stable set is contained in the core of the game (Jackson and Watts 2002, Klaus et al. 2010, Newton and Sawa 2015).

The key difference between marriage problems and assignment games is that the former are framed in terms of nontransferable (usually ordinal) utility, whereas in the latter each potential match has a transferable 'value'. The core of the assignment game consists of outcomes such that the matching is optimal and the allocation is pairwise stable. Generically, the optimal matching is unique and the allocations supporting it infinite. On the face of it one might suppose that the known results for marriage games would carry over easily to assignment games but this is not the case. The difficulty is that in marriage games (and roommate games) a payoff-improving deviation is determined by the players' current matches and their preferences, whereas in an assignment game it is determined by their matches, the value created by these matches, and by how they currently split the value of the matches. Thus the core of the assignment game tends to be significantly more constrained and paths to the core are harder to find than in the marriage game.

The contribution of the present work is to demonstrate a simple completely uncoupled adjustment process that has strong selection properties for assignment games. Using a proof technique introduced by Newton and Sawa (2015) (the one-period deviation principle), we show that the stochastically stable solutions of our process lie in a subset of the core of the assignment game which have a natural equity interpretation.

3.3 Matching markets with transferable utility

In this section we shall introduce the conceptual framework for analyzing matching markets with transferable utility; in the next section we introduce the learning process itself.

3.3.1 The assignment game

The population $V = \mathcal{F} \cup \mathcal{W}$ consists of firms $\mathcal{F} = \{f_1, ..., f_m\}$ and workers $\mathcal{W} = \{w_1, ..., w_n\}$.[5] They interact by making bids and offers to randomly encountered potential partners. We assume matches form only if these bids are mutually profitable for both agents.

Willingness to pay. Each firm i has a *willingness to pay*, $r_i^+(j) \geq 0$, for being matched with worker j.

Willingness to accept. Each worker j has a *willingness to accept*, $r_j^-(i) \geq 0$, for being matched with firm i.

We assume that these numbers are specific to the agents and are not known to the other market participants or to a central market authority.

Match value. Assume that utility is linear and separable in money. The *value* of a match $(i, j) \in \mathcal{F} \times \mathcal{W}$ is the potential surplus

$$w_{ij} = (r_i^+(j) - r_j^-(i))_+. \tag{3.1}$$

It will be convenient to assume that all values $r_i^+(j)$, $r_j^-(i)$, and w_{ij} can be

[5]The two sides of the market could also, for example, represent buyers and sellers, or men and women in a (monetized) marriage market.

expressed as multiples of some minimal unit of currency δ, for example, 'dollars'.

We shall introduce time at this stage to consistently develop our notation. Let $t = 0, 1, 2, ...$ be the time periods.

Assignment. For all pairs of agents $(i, j) \in \mathcal{F} \times \mathcal{W}$, let $m_{ij}^t \in \{0, 1\}$.

$$\text{If } (i,j) \text{ is} \begin{cases} \textit{matched} & \text{then} \quad m_{ij}^t = 1, \\ \textit{unmatched} & \text{then} \quad m_{ij}^t = 0. \end{cases} \tag{3.2}$$

If for a given agent $i \in V$ there exists j such that $m_{ij}^t = 1$ we shall refer to that agent as *matched*; otherwise i is *single*. An assignment $M^t = (m_{ij}^t)_{i \in \mathcal{F}, j \in \mathcal{W}}$ is such that if $m_{ij}^t = 1$ for some (i, j), then $m_{ik}^t = 0$ for all $k \neq j$ and $m_{lj}^t = 0$ for all $l \neq i$.

Matching market. The *matching market* is described by $[\mathcal{F}, \mathcal{W}, W, M]$:

- $\mathcal{F} = \{f_1, ..., f_m\}$ is the set of m firms (or men or buyers),
- $\mathcal{W} = \{w_1, ..., w_n\}$ is the set of n workers (or women or sellers),
- $W = \begin{pmatrix} w_{11} & \dots & w_{1n} \\ \vdots & w_{ij} & \vdots \\ w_{m1} & \dots & w_{mn} \end{pmatrix}$ is the matrix of match values.
- $M = \begin{pmatrix} m_{11} & \dots & m_{1n} \\ \vdots & m_{ij} & \vdots \\ m_{m1} & \dots & m_{mn} \end{pmatrix}$ is the assignment matrix with 0/1 values and row/column sums at most one.

The set of all possible assignments is denoted by $\mathcal{M}$.

Note, that the game at hand is a cooperative game:

Cooperative assignment game. Given $[\mathcal{F}, \mathcal{W}, W]$, the *cooperative assignment game* $G(\nu, V)$ is defined as follows. Let $\nu : S \subseteq V \to \mathbb{R}$ such that

- $\nu(i) = v(\varnothing) = 0$ for all singletons $i \in V$,
- $\nu(S) = w_{ij}$ for all $S = (i, j)$ such that $i \in \mathcal{F}$ and $j \in \mathcal{W}$,
- $\nu(S) = \max_{P \in \text{perm}(S)}\{\nu(i_{P(1)}, j_{P(1)}) + ... + \nu(i_{P(k)}, j_{P(k)})\}$
  $\forall S = (i_1, j_1, i_2, j_2, ..., i_{k^{\mathcal{F}}}, j_{k^{\mathcal{W}}}) \subseteq V$ and $k = \min\{k^{\mathcal{F}}, k^{\mathcal{W}}\}$.

The number $v(V)$ specifies the value of an optimal assignment.

3.3.2 Dynamic components

Aspiration level. At the end of any period t, a player has an *aspiration level*, d_i^t, which determines the minimal payoff at which he is willing to be matched. Let $\mathbf{d}^t = \{d_i^t\}_{i \in \mathcal{F} \cup \mathcal{W}}$.

Bids. In any period t, one pair of players is drawn at random and they make bids for each other. We assume that the two players' bids are such that the resulting payoff to each player is at least equal to his aspiration level, and with positive probability is exactly equal to his aspiration level.

Formally, firm $i \in \mathcal{F}$ encounters $j \in \mathcal{W}$ and submits a random bid $b_i^t = p_{ij}^t$, where p_{ij}^t is the maximal amount i is currently willing to pay if matched with j. Similarly, worker $j \in W$ submits $b_j^t = q_{ij}^t$, where q_{ij}^t is the minimal amount j is currently willing to accept if matched with i. A bid is separable into two components; the current (deterministic) aspiration level and a random variable that represents an exogenous shock to the agent's aspiration level. Specifically let P_{ij}^t, Q_{ij}^t be independent random variables that take values in $\delta \cdot \mathbb{N}_0$ where 0 has positive probability.[6] We thus have for all i, j

$$p_{ij}^t = (r_i^+(j) - d_i^{t-1}) - P_{ij}^t \qquad \text{and} \qquad q_{ij}^t = (r_j^-(i) + d_j^{t-1}) + Q_{ij}^t \tag{3.3}$$

[6]Note that $\mathbb{P}[P_{ij}^t = 0] > 0$ and $\mathbb{P}[Q_{ij}^t = 0] > 0$ are reasonable assumptions, since we can adjust $r_i^+(j)$ and $r_j^-(i)$ in order for it to hold. This would alter the underlying game but then allow us to proceed as suggested.

Consider, for example, worker j's bid for firm i. The amount $r_j^-(i)$ is the minimum that j would ever accept to be matched with i, while d_j^{t-1} is his previous aspiration level over and above the minimum. Thus Q_{ij}^t is j's attempt to get even more in the current period. Note that if the random variable is zero, the agent bids exactly according to his current aspiration level.

Prices. When i is matched with j they trade at a unique *price*, π_{ij}^t.

Payoffs. Given $[M^t, \mathbf{d^t}]$ the *payoff* to firm i / worker j is

$$\phi_i^t = \begin{cases} r_i^+(j) - \pi_{ij}^t & \text{if } i \text{ is matched to } j, \\ 0 & \text{if } i \text{ is single.} \end{cases} \tag{3.4}$$

$$\phi_j^t = \begin{cases} \pi_{ij}^t - r_j^-(i) & \text{if } j \text{ is matched to } i, \\ 0 & \text{if } j \text{ is single.} \end{cases} \tag{3.5}$$

Note that players' payoffs can be deduced from the aspiration levels and the assignment matrix.

Profitability. A pair of bids (p_{ij}^t, q_{ij}^t) is *profitable* if both players, in expectation, receive a higher payoff if the match is formed.

Note that if two players' bids are at their aspiration levels and $p_{ij}^t = q_{ij}^t$ they are only profitable if both players are currently single. Also note that a pair of players (i, j) with $w_{ij} = 0$ will never match.

Re-match. At each moment in time, a pair (i, j) that randomly encounters each other matches if their bids are profitable. The resulting price, π_{ij}^t, is set anywhere between q_{ij}^t and p_{ij}^t. (Details about how players are activated are specified in the next section.)

To summarize, when a new match forms that is profitable, both agents receive a higher payoff in expectation due to the full support of the resulting price.[7]

States. The *state* at the end of period t is given by $Z^t = [M^t, \mathbf{d}^t]$ where $M^t \in \mathcal{M}$ is an assignment and $\mathbf{d}^t$ is the aspiration level vector. Denote the set of all states by Ω.

3.3.3 Solution concepts

Optimality. An assignment M is *optimal* if $\sum_{(i,j)\in\mathcal{F}\times\mathcal{W}} a_{ij} \cdot w_{ij} = v(V)$.

Pairwise stability. An aspiration level vector $\mathbf{d}^t$ is *pairwise stable* if $\forall i, j$ and $a^t_{ij} = 1$,

$$r_i^+(j) - d_i^t = r_j^-(i) + d_j^t, \tag{3.6}$$

and $r_{i'}^+(j) - d_{i'}^t \leq r_j^-(i') + d_j^t$ for every alternative firm $i' \in \mathcal{F}$ with $i' \neq i$ and $r_j^-(i') + d_{j'}^t \geq r_{i'}^+(j) - d_i^t$ for every alternative worker $j' \in \mathcal{W}$ with $j' \neq j$.

Core (Shapley and Shubik 1972). The *core* of any assignment game is always non-empty and consists of the set $\mathbf{C} \subseteq \Omega$ of all states Z such that M is an optimal assignment and $\mathbf{d}$ is pairwise stable.

Subsequent literature has investigated the structure of the assignment game core, which turns out to be very rich.[8] In order to investigate the constraints of pairwise stability in more detail the concept of 'payoff excess' will be useful:

Excess. Given state Z^t, the *excess* for a player i who is matched with j is

$$e_i^t = \phi_i^t - \max_{k \neq j}(w_{ik} - \phi_k^t)_+. \tag{3.7}$$

[7]In this sense any alternative match that may block a current assignment because it is profitable (as defined earlier) is a strict blocking pair.

[8]See, for example, Roth and Sotomayor (1992), Balinski and Gale (1987), Sotomayor (2003).

The excess for player i describes the gap to his next-best alternative, that is, the smallest amount he would have to give up in order to profitably match with some other player $k \neq j$. If a player has negative excess, pairwise stability is violated. In a core allocation, therefore, all players have nonnegative excess. For the analysis of absorbing core states, note that the excess in payoff can be equivalently expressed in terms of the excess in aspiration level. This is the case since in absorbing core states aspiration levels are directly deducible from payoffs.

Minimal excess. Given state Z^t, the *minimal excess* is

$$e^t_{\min}(Z^t) = \min_{i:\, i \text{ matched}} e^t_i. \tag{3.8}$$

Based on the minimal excess of a state, we can define the kernel (Davis and Maschler 1965). For assignment games, the kernel coincides with the solution concept proposed by Rochford (1984), which generalizes a pairwise equal split solution à la Nash (1950a).

Kernel (Davis and Maschler 1965, Rochford 1984). The *kernel* $\mathbf{K}$ of an assignment game is the set of states such that the matching is optimal and for all matched pairs (i, j),

$$e^t_i =_\delta e^t_j, \tag{3.9}$$

where $=_\delta$ means "equality up to δ". (This is necessary given that we operate on the discrete grid.)

Given Z^t, extend the definition of excess to *coalitional excess* for coalition $S \subseteq V$; $e^t(S) = \sum_{i \in S} \phi^t_i - v(S)$. Now let $\mathbf{E}(\phi^t) \in \mathbb{R}^{m+n}$ be the vector of coalitional excesses for all $S \subseteq V$, ordered from smallest to largest. Say $\mathbf{E}(\phi)$ is lexicographically larger than $\mathbf{E}(\phi')$ for some k, if $\mathbf{E}_i(\phi) = \mathbf{E}_i(\phi')$ for all $i < k$

and $\mathbf{E}_k(\phi) < \mathbf{E}_k(\phi')$.[9]

Nucleolus (Schmeidler 1969). The *nucleolus* $\mathbf{N}$ of the assignment game is the unique solution that minimizes the lexicographic measure. (See also Huberman 1980, Solymosi and Raghavan 1994.)

For an analysis of the welfare properties and of the links between the kernel and the nucleolus of the assignment game see Nunez (2004) and Llerena et al. (2015).

Least core (Maschler et al. 1979). The *least core* $\mathbf{L}$ of an assignment game is the set of states Z such that the matching is optimal and the minimum excess is maximized, that is,

$$e_{\min}(Z) = \max_{Z' \in \mathbf{C}} e_{\min}(Z'). \tag{3.10}$$

Note that our definition of excess (equations (3.7) and (3.8)) applies to essential coalitions only (that is, for the case of the assignment game, to two-player coalitions involving exactly one agent from each market side). Hence, the least core generalizes the nucleolus of the assignment game in the following sense. Starting with the nucleolus, select any player with minimum excess (according to equation (3.7)): the least core contains all outcomes with a minimum excess that is not smaller.[10]

The following inclusions are known for the assignment game:[11]

$$\mathbf{N} \in (\mathbf{K} \cap \mathbf{L}),\ \ \mathbf{K} \subseteq \mathbf{C},\ \ \mathbf{L} \subseteq \mathbf{C}. \tag{3.11}$$

[9]Note that the excess for coalitions, $e^t(S)$, is usually defined with a reversed sign. In order to make it more concurrent in light of definition (3.7) we chose to reverse the sign.

[10]See Shapley and Shubik (1963, 1966) for the underlying idea of the least core, the *strong ε-core*. See Maschler et al. (1979), Driessen (1999), Llerena and Nunez (2011) for geometric interpretations of these concepts.

[11]$\mathbf{N} \in \mathbf{K}$ is shown by Schmeidler (1969) for general cooperative games. Similarly $\mathbf{N} \in \mathbf{L}$ is shown by Maschler et al. (1979). Driessen (1998) shows for the assignment game that $\mathbf{K} \subseteq \mathbf{C}$. $\mathbf{L} \subseteq \mathbf{C}$ follows directly from the definitions.

3.4 Evolving play

A fixed population of agents, $V = \mathcal{F} \cup \mathcal{W}$, plays the assignment game $G(v, V)$. Repeatedly, a randomly activated agent encounters another agent, they make bids for each other and match if profitable. The distinct times at which one agent becomes active will be called *periods*. Agents are activated by independent Poisson clocks.[12] Suppose that an active agent randomly encounters one agent from the other side of the market drawn from a distribution with full support. The two players enter a new match if their match is profitable, which they can see from their current bids, offers and their own payoffs. If the two players are already matched with each other, they remain so.

3.4.1 Behavioral dynamics

The essential steps and features of the learning process are as follows. At the start of period $t + 1$:

1. The activated agent, i, makes a random encounter, j.

2a. If the encounter is profitable given their current bids and assignment, the pair matches.

2b. If the match is not profitable, both agents return to their previous matches (or remain single).

3a. If a new match (i, j) forms, the price is set anywhere between bid and offer. The aspiration levels of i and j are set to equal their realized payoffs.

[12] The Poisson clocks' arrival rates may depend on the agents' themselves or on their position in the game. Single agents, for example, may be activated faster than matched agents.

3b. If no new match is formed, the active agent, if he was previously matched, keeps his previous aspiration level and stays with his previous partner. If he was previously single, he remains single and lowers his aspiration level with positive probability.

Our rules have antecedents in the psychology literature (Thorndike 1898, Hoppe 1931, Estes 1950, Bush and Mosteller 1955, Herrnstein 1961). To the best of our knowledge, however, such a framework has not previously been used in the study of matching markets in cooperative games. The approach seems especially well-suited to modeling behavior in large decentralized assignment markets, where agents have little information about the overall game and about the identity of the other market participants. Following aspiration adjustment theory (Sauermann and Selten 1962, Selten 1998) and related bargaining experiments on directional and reinforcement learning (e.g., Tietz and Weber 1972, Roth and Erev 1995), we shall assume a simple directional learning model: matched agents occasionally experiment with higher offers if on the sell-side (or lower bids if on the buy-side), while single agents, in the hope of attracting partners, lower their offers if on the sell-side (or increase their bids if on the buy-side).

We shall now describe the process in more detail, distinguishing the cases where the active agent is currently *matched* or *single*. Let Z^t be the state at the end of period t (and the beginning of period $t+1$), and let $i \in F$ be the unique *active* agent which for ease of exposition we assume to be a firm.

I. The active agent is currently matched and meets j

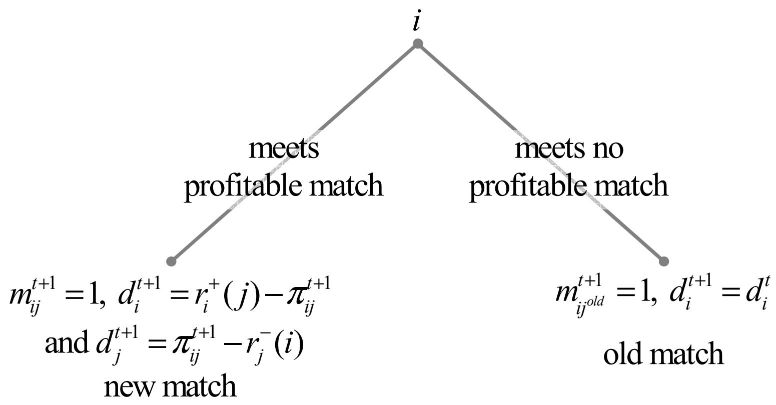
If i, j are profitable (given their current aspiration levels) they match. As a result, i's former partner is now single (and so is j's former partner if j was matched in period t). The price governing the new match, π_{ij}^{t+1}, is randomly set between p_{ij}^{t+1} and q_{ij}^{t+1}.

At the end of period $t+1$, the aspiration levels of the newly matched pair (i,j) are adjusted according to their newly realized payoffs:

$$d_i^{t+1} = r_i^+(j) - \pi_{ij}^{t+1} \quad \text{and} \quad d_j^{t+1} = \pi_{ij}^{t+1} - r_j^-(i). \tag{3.12}$$

All other aspiration levels and matches remain fixed. If i,j are not profitable, i remains matched with his previous partner and keeps his previous aspiration level. See Figure 3.1 for an illustration.

Figure 3.1: Transition diagram for active, matched agent (period $t+1$).



II. The active agent is currently single and meets j

If i,j are profitable (given their current aspiration levels) they match. As a result, j's former partner is now single if j was matched in period t. The price governing the new match, π_{ij}^{t+1}, is randomly set between p_{ij}^{t+1} and q_{ij}^{t+1}.

At the end of period $t+1$, the aspiration levels of the newly matched pair (i,j) are adjusted to equal their newly realized payoffs:

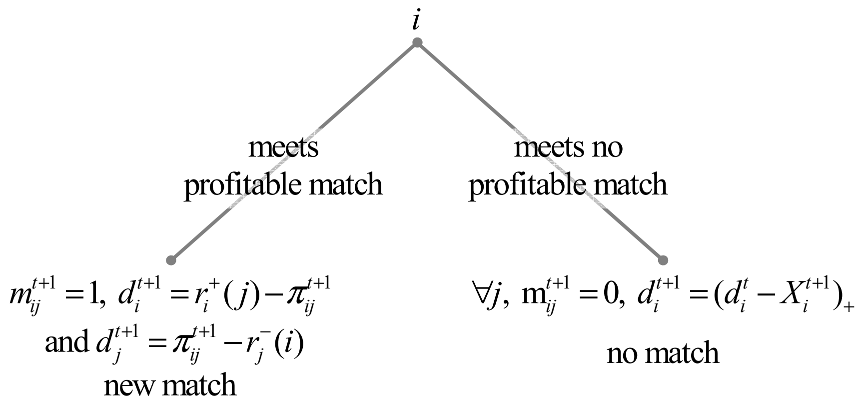
$$d_i^{t+1} = r_i^+(j) - \pi_{ij}^{t+1} \quad \text{and} \quad d_j^{t+1} = \pi_{ij}^{t+1} - r_j^-(i). \tag{3.13}$$

All other aspiration levels and matches remain as before. If i, j are not profitable, i remains single and, with positive probability, reduces his aspiration level,

$$d_i^{t+1} = (d_i^t - X_i^{t+1})_+, \tag{3.14}$$

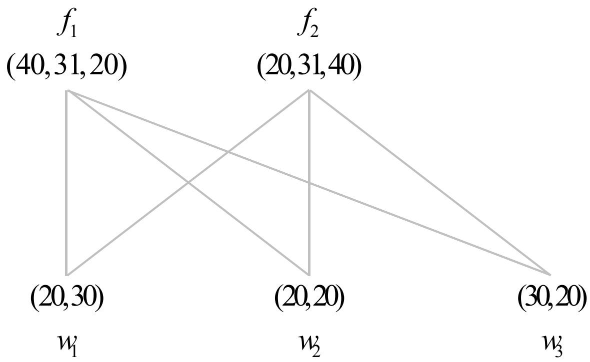
where X_i^{t+1} is an independent random variable taking values in $\delta \cdot \mathbb{N}_0$, and δ occurs with positive probability.[13] See Figure 3.2 for an illustration.

Figure 3.2: Transition diagram for active, single agent (period $t+1$).



3.4.2 Example

Let $V = \mathcal{F} \cup \mathcal{W} = \{f_1, f_2\} \cup \{w_1, w_2, w_3\}$, $r_1^+(j) = (40, 31, 20)$ and $r_2^+(j) = (20, 31, 40)$ for $j = 1, 2, 3$, and $r_1^-(i) = (20, 30)$, $r_2^-(i) = (20, 20)$ and $r_3^-(i) = (30, 20)$ for $i = 1, 2$.



[13]Note that X_i^{t+1} may depend on time.

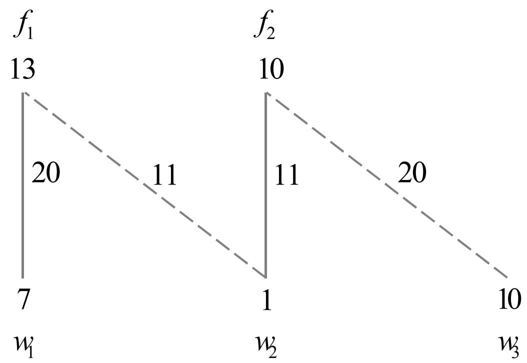
Then one can compute the match values: $w_{11} = w_{23} = 20$, $w_{12} = w_{22} = 11$, and $w_{ij} = 0$ for all other pairs (i, j). Let $\delta = 1$.

period t: *Current state*

Suppose that, at the end of some period t, (f_1, w_1) and (f_2, w_2) are matched and w_3 is single.

The current aspiration level is shown next to the name of that agent, and the values w_{ij} are shown next to the edges (if positive). Bids will be shown to the right of the aspiration level. Solid edges indicate matched pairs, and dashed edges indicate unmatched pairs. (Edges with value zero are not shown.) Note that no player can see the bids or the status of the players on the other side of the market.

Note that some matches can never occur. For example f_1 is never willing to pay more than 20 for w_3, but w_3 would only accept a price above 30 from f_1.

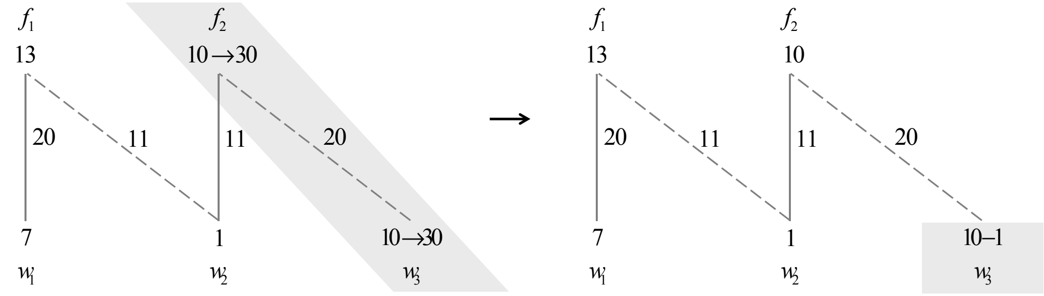


Note that the aspiration levels satisfy $d_i^t + d_j^t \geq w_{ij}$ for all i and j, but the assignment is not optimal (firm 2 should match with worker 3).

period $t+1$: *Activation of single agent w_3 and encounter of f_2*

w_3's current aspiration level is too high in order to be profitable with f_2. Hence, independent of the specific bid he makes, he remains single and, with positive probability, reduces his aspiration level by 1.

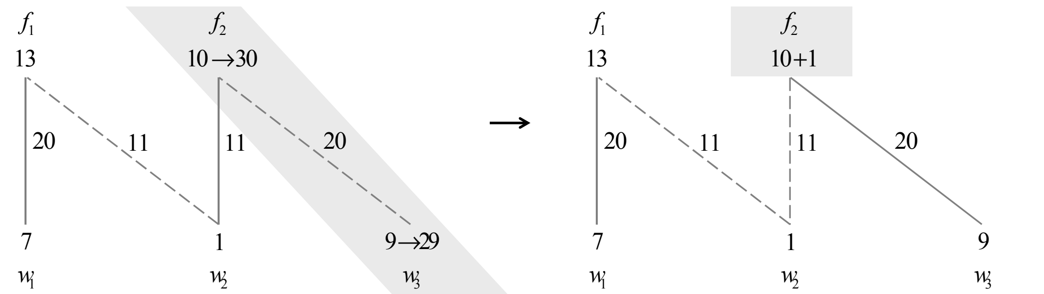
w_3 encounters f_2 $\quad$ w_3 reduces aspiration level



period $t+2$**:** *Activation of matched agent* f_2 *and encounter of* w_3

f_2 and w_3 are profitable. With positive probability f_2 bids 30 for w_3 and w_3 bids 29 for f_2 (hence the match is profitable), and the match forms. The price is set at random to either 29 such that f_2 raises his aspiration level by one unit (11) and w_3 keeps his aspiration level (9), or to 30 such that f_2 keeps his aspiration level (10) and w_3 raises his aspiration level by one unit (10). (Thus in expectation the agents get a higher payoff than before.)

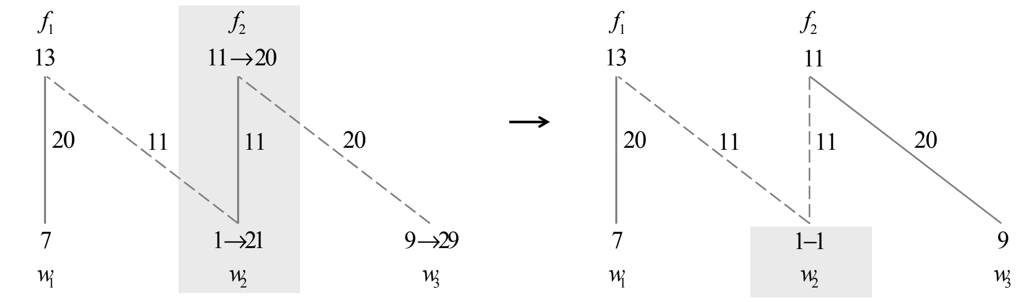
f_2 encounters w_3 $\quad$ Successful match; f_2 increases aspiration level



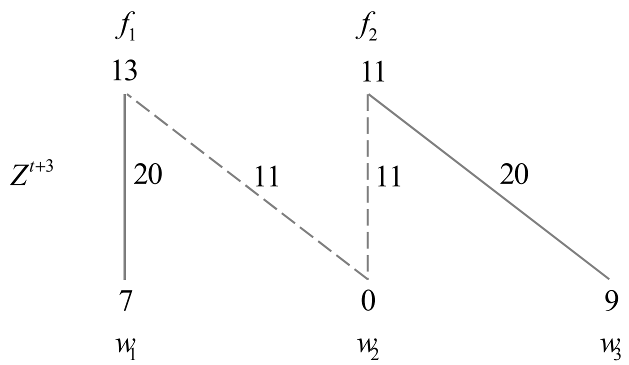
period $t+3$**:** *Activation of single agent* w_2 *and encounter of* f_2

w_2's current aspiration level is too high in the sense that he has no profitable matches and thus in particular is not profitable with f_2. Hence he remains single and, with positive probability, reduces his aspiration level by 1.

w_2 encounters f_2 $\qquad\qquad$ w_2 reduces aspiration level



The resulting state is in the core:[14]



3.5 Core stability – absorbing states of the unperturbed process

Recall that a state Z^t is defined by an assignment M^t and aspiration levels $\mathbf{d}^t$ that jointly determine the payoffs. $\mathbf{C}$ is the set of core states; let $\mathbf{C}_0$ be the set of core states such that singles' aspiration levels are zero.

Theorem 12. *Given an assignment game $G(v, V)$, from any initial state $Z^0 = [M^0, \mathbf{d}^0] \in \Omega$, the process is absorbed into the core in finite time with probability 1. The set of absorbing states consists of $\mathbf{C}_0$. Further, starting from $\mathbf{d}^0 = 0$ any absorbing state is attainable.*

[14]Note that the states Z^{t+2} and Z^{t+3} are both in the core, but Z^{t+3} is absorbing whereas Z^{t+2} is not.

Throughout the proof we shall omit the time superscript since the process is time-homogeneous. The general idea of the proof is to show a particular path leading into the core which has positive probability. The proof uses integer programming arguments (Kuhn 1955, Balinski 1965) but no single authority 'solves' an integer programming problem. It will simplify the argument to restrict our attention to a particular class of paths with the property that the realizations of the random variables P_{ij}^t, Q_{ij}^t are always 0 and the realizations of X_i^t are always δ. P_{ij}^t, Q_{ij}^t determine the gaps between the bids and the aspiration levels, and X_i^t determines the reduction of the aspiration level by a single agent. One obtains from equation (3.3) for the bids:

$$\text{for all } i, j, \qquad p_{ij}^t = r_i^+(j) - d_i^{t-1} \qquad \text{and} \qquad q_{ij}^t = r_j^-(i) + d_j^{t-1} \tag{3.15}$$

Recall that any two agents encounter each other in any period with positive probability. It shall be understood in the proof that the relevant agents in any period encounter each other. Jointly with equation (3.3), we can then say that a pair of aspiration levels (d_i^t, d_j^t) is *profitable* if

$$\text{either} \quad d_i^t + d_j^t < w_{ij} \quad \text{or} \quad d_i^t + d_j^t = w_{ij} \ \text{ and both } i \text{ and } j \text{ are single.} \tag{3.16}$$

Restricting attention to this particular class of paths will permit a more transparent analysis of the transitions, which we can describe solely in terms of the aspiration levels.

We shall proceed by establishing the following two claims.

Claim 1. There is a positive probability path to aspiration levels $\mathbf{d}$ such that $d_i + d_j \geq w_{ij}$ for all i, j and such that, for every i, either there exists a j such that $d_i + d_j = w_{ij}$ or else $d_i = 0$.

Any aspiration levels satisfying Claim 1 will be called *good*. Note that, even if aspiration levels are good, the assignment does not need to be optimal and not every agent with a positive aspiration level needs to be matched. (See the period-t example in the preceding section.)

Claim 2. Starting at any state with good aspiration levels, there is a positive probability path to a pair $(M, \mathbf{d})$ where $\mathbf{d}$ is good, M is optimal, and all singles' aspiration levels are zero.[15]

Proof of Claim 1.

Case 1. Suppose the aspiration levels $\mathbf{d}$ are such that $d_i + d_j < w_{ij}$ for some i, j. Note that this implies that i and j are not matched with each other since otherwise the entire surplus is allocated and $d_i + d_j = w_{ij}$. With positive probability, either i or j is activated and i and j become matched. The new aspiration levels are set equal to the new payoffs. Thus the sum of the aspiration levels is equal to the match value w_{ij}. Therefore, there is a positive probability path along which $\mathbf{d}$ increases monotonically until $d_i + d_j \geq w_{ij}$ for all i, j.

Case 2. Suppose the aspiration levels $\mathbf{d}$ are such that $d_i + d_j \geq w_{ij}$ for all i, j.

We can suppose that there exists a single agent i with $d_i > 0$ and $d_i + d_j > w_{ij}$ for all j, else we are done. With positive probability, i is activated. Since no profitable match exists, he lowers his aspiration level by δ. In this manner, a suitable path can be constructed along which $\mathbf{d}$ decreases monotonically until the aspiration levels are good. Note that at the end of such a path, the assignment does not need to be optimal and not every agent with a positive aspiration level needs to be matched. (See the period-t example in the preceding section.) ☺

[15]Note that this claim describes an absorbing state in the core. It may well be that the core is reached while a single's aspiration level is more than zero. The latter state, however, is transient and will converge to the corresponding absorbing state.

Proof of Claim 2.

Suppose that the state $(M, \mathbf{d})$ satisfies Claim 1 ($\mathbf{d}$ is good) and that some single exists whose aspiration level is positive. (If no such single exists, the assignment is optimal and we have reached a core state.) Starting at any such state, we show that, within a bounded number of periods and with positive probability (bounded below), one of the following holds:

The aspiration levels are good, the number of single agents with positive aspiration level decreases, and the sum of the aspiration levels remains constant. (3.17)

The aspiration levels are good, the sum of the aspiration levels decreases by $\delta > 0$, and the number of single agents with a positive aspiration level does not increase. (3.18)

In general, say an edge is *tight* if $d_i + d_j = w_{ij}$ and *loose* if $d_i + d_j = w_{ij} - \delta$. Define a *maximal alternating path* P to be a path that starts at a single player with positive aspiration level, and that alternates between unmatched tight edges and matched tight edges such that it can not be extended (hence maximal). Note that, for every single with a positive aspiration level, at least one maximal alternating path exists. Figure 3.3 (left panel) illustrates a maximal alternating path starting at f_1. Unmatched tight edges are indicated by dashed lines, matched tight edges by solid lines and loose edges by dotted lines.

Without loss of generality, let f_1 be a single firm with positive aspiration level.

Case 1. Starting at f_1, there exists a maximal alternating path P of odd length.

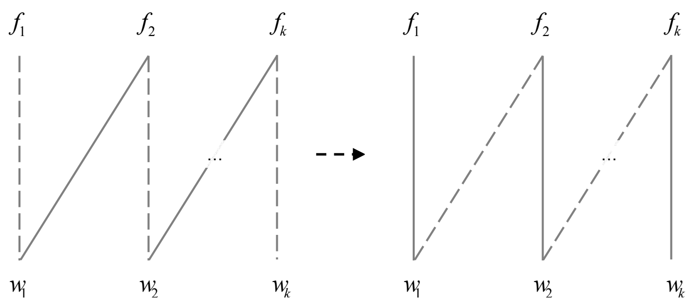
Case 1a. All firms on the path have a positive aspiration level.

We shall demonstrate a sequence of adjustments leading to a state as in (3.17).

Let $P = (f_1, w_1, f_2, w_2, ..., w_{k-1}, f_k, w_k)$. Note that, since the path is maximal and of odd length, w_k must be single. With positive probability, f_1 is activated. Since no profitable match exists, he lowers his aspiration level by δ. With positive probability, f_1 is activated again next period, he snags w_1 and with positive probability he receives the residual δ. At this point the aspiration levels are unchanged but f_2 is now single. With positive probability, f_2 is activated. Since no profitable match exists, he lowers his aspiration level by δ. With positive probability, f_2 is activated again next period, he snags w_2 and with positive probability he receives the residual δ. Within a finite number of periods a state is reached where all players on P are matched and the aspiration levels are as before. (Note that f_k is matched with w_k without a previous reduction by f_k since w_k is single and thus their bids are profitable.)

In summary, the number of matched agents has increased by two and the number of single agents with positive aspiration level has decreased by at least one. The aspiration levels did not change, hence they are still good. (See Figure 3.3 for an illustration.)

Figure 3.3: Transition diagram for Case 1a.



Case 1b. At least one firm on the path has aspiration level zero.

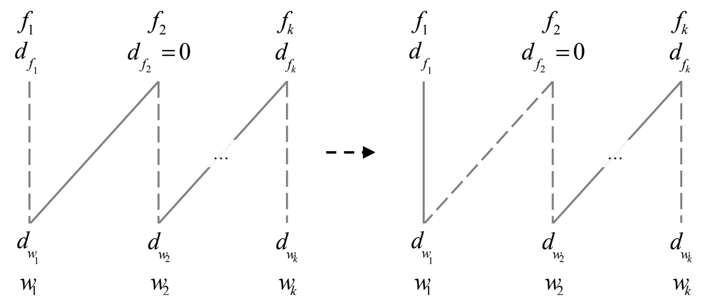
We shall demonstrate a sequence of adjustments leading to a state as in (3.17).

Let $P = (f_1, w_1, f_2, w_2, ..., w_{k-1}, f_k, w_k)$. There exists a firm $f_i \in P$ with current aspiration level zero (f_2 in the illustration), hence no further reduction by f_i

can occur. (If multiple firms on P have aspiration level zero, let f_i be the first such firm on the path.) Apply the same sequence of transitions as in Case 1a up to firm f_i. At the end of this sequence the aspiration levels are as before. Once f_{i-1} snags w_{i-1}, f_i becomes single and his aspiration level is still zero.

In summary, the number of single agents with a positive aspiration level has decreased by one because f_1 is no longer single and the new single agent f_i has aspiration level zero. The aspiration levels did not change, hence they are still good. (See Figure 3.4 for an illustration.)

Figure 3.4: Transition diagram for Case 1b.



Case 2. Starting at f_1, *all* maximal alternating paths are of even length.

Case 2a. All firms on all maximal alternating paths starting at f_1 have a positive aspiration level.

We shall demonstrate a sequence of adjustments leading to a state as in (3.18).

Since aspiration levels will have changed by the end of the sequence of transitions, it does not suffice to only consider players along one maximal alternating path (for otherwise the aspiration levels will no longer be good). Instead, we need to consider *the union* of alternating paths (which are sets of edges) starting at f_1. Note that this union is a connected graph, say G_{f_1}. (Players may be part of multiple maximal alternating paths starting at f_1.) Now, fix a subgraph $T_{f_1} \subseteq G_{f_1}$ that includes all firms and workers, is connected, has no cycles and contains all of the matched edges. T_{f_1} is a *spanning tree* of G_{f_1} and thus always exists. Note that T_{f_1} only has alternating paths (between unmatched tight and matched tight edges) and all maximal alternating paths on this tree which start at f_1 contain an even number of edges. Thus there is one more firm on this graph than there are workers. (Think of f_1 as that extra firm.)

We shall describe a sequence of transitions along this tree such that, at its end, all firms on the graph have reduced their aspiration level by δ and all workers have increased their aspiration level by δ. We shall label firms and workers such that, for $i \geq 2$, f_i is at distance $2i - 2$ from f_1 on the tree T_{f_1}, and, for $i \geq 1$, w_i is at distance $2i - 1$ from f_i. Note that this implies that, for two agents with label i and j and $i < j$, the former agent is closer to f_1 than the latter. Let k be the maximal i such that f_i is on T_{f_1}.

With positive probability f_1 is activated. Since no profitable match exists, f_1 lowers his aspiration level by δ. Hence, all previously tight edges starting at f_1 are now loose.

We shall describe a sequence of transitions which will tighten a loose edge along the tree T_{f_1} by making another edge on the tree loose. This new loose edge is further away from f_1 on the tree. At the end of this sequence the matching will not have changed and the sum of aspiration levels will have remained fixed.

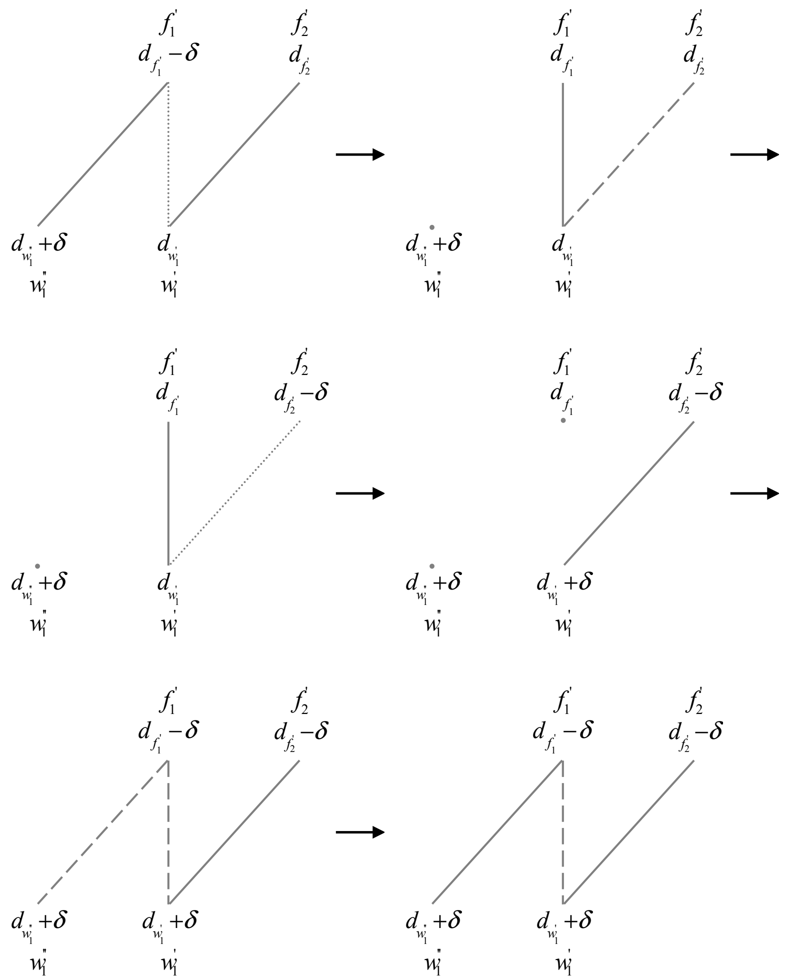
Consider one such loose edge starting at $f_i \in T_{f_1}$. The loose edge from f_i is shared with a worker, say w_i. Since all maximal alternating paths starting at f_1 are of even length, the worker has to be matched to a firm, say f_{i+1}. With positive probability, w_i is activated, matches with f_i, and f_i receives the residual δ. (Such a transition occurs with strictly positive probability, whether or not f_i is matched, because the sum of aspiration levels is strictly below the match value of (w_i, f_i).) Note that f_{i+1} and, for $i \neq 1$, f_i's previous partner, w_{i-1}, are now single. With positive probability, f_{i+1} is activated and meets

w_i again who is now matched with f_i. Since the aspiration levels of both players have not changed, their match is not profitable, and thus f_{i+1} lowers his aspiration level by δ. (Note that this reduction is possible because all firms on any maximal alternating path starting at f_1 have aspiration level at least δ.) With positive probability, f_{i+1} is activated again, matches with w_i, and w_i receives the residual δ. Finally, with positive probability, f_i is activated. Since no profitable match exists, he lowers his aspiration level by δ. For $i \neq 1$, f_i was previously matched, then f_i is activated with positive probability again in the next period and matches with the single w_{i-1} (there is no additional surplus to be split).

At the end of the sequence described above the assignment is the same as at the beginning. Moreover, w_i's aspiration level went up by δ while f_{i+1}'s aspiration level went down by δ and all other aspiration levels stayed the same. The originally loose edge between f_i and w_i is now tight. If f_{i+1} is not the last firm on its maximal alternating paths on T_{f_1}, there will be a new loose edge on T_{f_1} between f_{i+1} and workers with label $i+1$ (at distance $2i+1$). Note that there may be other loose edges not on T_{f_1}. These we shall not consider in our construction since they disappear at its end. (See Figure 3.5 for an illustration.)

We repeat the latter construction on T_{f_1} for $f_i = f_1$ until all loose edges at f_1 have been eliminated (*iteration 1*). This may create new loose edges on T_{f_1} between firms at distance 2 from f_1 (on T_{f_1}) and workers at distance 3 from f_1. Next we repeat the construction for all firms at distance 2 from f_1 on T_{f_1} (*iteration 2*). Note that the distance of loose edges to f_1 is increasing after each such iteration (recall that we consider the fixed spanning tree T_{f_1}). We thus keep iterating the construction for firms at distance $2i$ from f_1 for $i = 2, 3, \ldots, k-1$. The final iteration of our construction on a maximal alternating

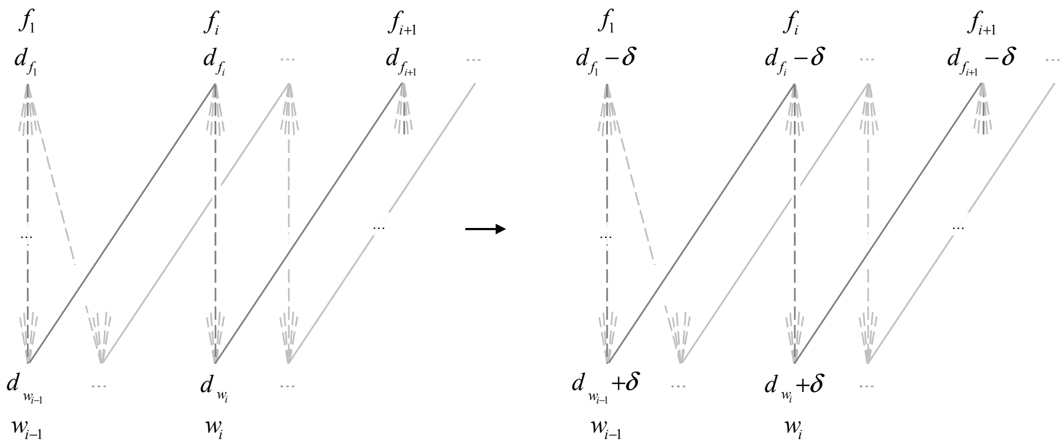
Figure 3.5: Detailed transitions for Case 2a.



path does not lead to new loose edges on T_{f_1} (for otherwise the last firm on the unique path from f_1 on T_{f_1} would previously have had a tight edge, contradicting the assumption that the alternating path is maximal.) Hence, in a finite number of periods, all firms on T_{f_1}, and thus on G_{f_1}, have reduced their aspiration level by δ, and all workers have increased their aspiration level by δ. Given that G_{f_1} contained *all* maximal alternating paths starting at f_1, and therefore all tight paths starting at f_1, there are no more loose edges connected to any firm or worker on G_{f_1}. This is the case for two reasons. First, all edges on G_{f_1} were tight to begin with and, after each firm reduced his aspiration level by δ and each worker increased his aspiration level by δ, they are again tight at the end of all iterations. Second, there cannot be loose edges with firms or workers outside G_{f_1} since they were not tight before, and thus one δ-reduction cannot make them loose.

In summary, aspiration level reductions outnumber aspiration level increases by one (namely by the δ-reduction by firm f_1), hence the sum of the aspiration levels has decreased. The number of single agents with a positive aspiration level has not increased. Moreover the aspiration levels are still good. (See Figure 3.6 for an illustration.)

Figure 3.6: Transition diagram for Case 2a.



Note that the δ-reductions may lead to new tight edges, resulting in new maximal alternating paths of odd or even lengths.

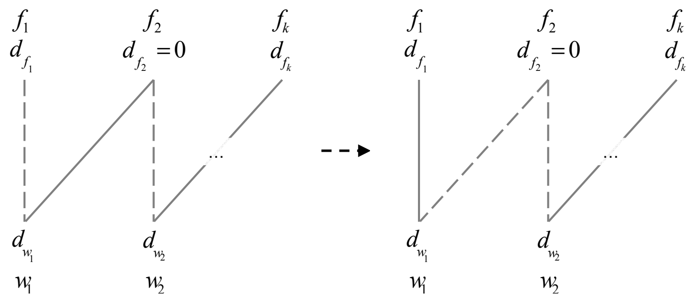
Case 2b. At least one firm on a maximal alternating paths starting at f_1 has aspiration level zero.

We shall demonstrate a sequence of adjustments leading to a state as in (3.17).

Let $P = (f_1, w_1, f_2, w_2, ..., w_{k-1}, f_k)$ be a maximal alternating path such that a firm has aspiration level zero. There exists a firm $f_i \in P$ with current aspiration level zero (f_2 in the illustration), hence no further reduction by f_i can occur. (If multiple firms on P have aspiration level zero, let f_i be the first such firm on the path.) With positive probability f_1 is activated. Since no profitable match exists, he lowers his aspiration level by δ. With positive probability, f_1 is activated again next period, he snags w_1 and with positive probability he receives the residual δ. Now f_2 is single. With positive probability f_2 is activated, lowers, snags w_2, and so forth. This sequence continues until f_i is reached, who is now single with aspiration level zero.

In summary, the number of single agents with a positive aspiration level has decreased. The aspiration levels did not change, hence they are still good. (See Figure 3.7 for an illustration.)

Figure 3.7: Transition diagram for Case 2b.



Let us summarize the argument. Starting in a state $[M, \mathbf{d}]$ with good aspiration levels $\mathbf{d}$, we successively (if any exist) eliminate the odd paths starting at firms/workers followed by the even paths starting at firms/workers, while maintaining good aspiration levels. This process must come to an end because at each iteration either the sum of aspiration levels decreases by δ and the number of single agents with positive aspiration levels stays fixed, or the sum of aspiration levels stays fixed and the number of single agents with positive aspiration levels decreases. The resulting state must be in the core and is absorbing because single agents cannot reduce their aspiration level further and no new matches can be formed. Since an aspiration level constitutes a lower bound on a player's bids we can conclude that the process Z^t is absorbed into the core in finite time with probability 1. Finally note that, starting from $\mathbf{d}^0 = 0$, we can trivially reach any state in $\mathbf{C}_0$. ☺

3.6 Core selection – stable states of the perturbed process

In this section, we investigate the effects of random perturbations to the adjustment process. Suppose that players occasionally experience shocks when in a match and that larger shocks are less likely than smaller shocks. The effect of such a shock is that a player receives more or less payoff than anticipated given the current price he agreed to with his partner. We shall formalize these perturbations and investigate the resulting selection of stochastically stable states as the probability of shocks becomes vanishingly small (see section 1.3). It turns out that the set of stochastically stable states is contained in the *least core*; moreover there are natural conditions under which it coincides with the least core.

Given a player i who is matched in period t. Suppose his unperturbed payoff ϕ_i^t is subject to a shock. Denote the new payoff by $\hat{\phi}_i^t$ and define:

$$\hat{\phi}_i^t = \begin{cases} \phi_i^t + \delta \cdot R_i^t \text{ with probability } 0.5, \\ \phi_i^t - \delta \cdot R_i^t \text{ with probability } 0.5, \end{cases} \tag{3.19}$$

where R_i^t is an independent geometric random variable with $\mathbb{P}[R_i^t = k] = \varepsilon^k \cdot (1 - \varepsilon)$ for all $k \in \mathbb{N}_0$.[16] Note that for $\varepsilon = 0$ the process is unperturbed.

The immediate result of a given shock is that players receive a different payoff

[16]For simplicity we propose this specific distribution. But note that any probability distribution can be assumed as long as there exists a parameter ε such that $\mathbb{P}[x + 1] = \varepsilon \cdot O(\mathbb{P}[x])$ for all $x \in \mathbb{N}_0$.

than anticipated. We shall assume that players update their aspiration levels to equal their new perturbed payoff if the payoff is positive and zero if it is negative. If, in a given match, one of the players experiences a shock that renders his payoff negative, then we assume that the match breaks up and both players become single. Note that, if the partnership remains matched, the price does not change. The latter implies that if i does not break up or re-match in the next period $(t+1)$ his unperturbed payoff in $t+1$ will again be ϕ_i^t, and the period-t shock then has no influence on subsequent state transitions.

3.6.1 Analysis

With the machinery of stochastic stability (see section 1.3) at hand we shall show that the stochastically stable states are contained in the least core. To establish this result we shall adapt a proof technique due to Newton and Sawa (2015) and show that the least core is the set of states which is most stable against one-shot deviations. We shall also provide conditions on the game under which the stochastically stable set is identical with the least core.

Recall from section 1.3 that a transition (possibly in multiple periods) $Z \to Z'$ a *least cost transition* if it exhibits the lowest order of resistance, that is, let $Z = Z_0, Z_1, \ldots, Z_k = Z'$ (k finite) be a path of one-period transitions from Z to Z'. Then a least cost transition minimizes $\sum_{l=0}^{k-1} r(Z_l, Z_{l+1})$ over all such paths. We shall say for a core state $Z \in \mathbf{C}$ that a transition out of the core is a *least cost transition* if it minimizes the resistance among all transitions from Z to any non-core state.

Recall that the least core consists of states that maximize the following term:

$$
\begin{aligned}
e^t_{\min} &= \min_{i:\, i \text{ matched}} \{\phi^t_i - \max_{j:\, m^t_{ij}=0} (w_{ij} - \phi^t_j)_+\} && (3.20)\\
&= \min\{ \underbrace{\min_{i,j:\, m^t_{ij}=0,\, i \text{ matched}} (\phi^t_i + \phi^t_j - w_{ij})}_{=:\, \text{case } A} \;;\; \underbrace{\min_{i:\, i \text{ matched}} \phi^t_i}_{=:\, \text{case } B} \} && (3.21)
\end{aligned}
$$

Case A holds if the minimal cost deviation is such that two players who are currently not matched experience shocks such that their match becomes profitable. Case B, on the other hand, is the case where the minimal cost deviation is such that a matched agent experiences a shock that leads to a negative payoff and thus to him breaking up his relationship.

Given two states Z and Z^*, let the *distance* between them be

$$D(Z, Z^*) = \sum_{i \in \mathcal{F} \cup \mathcal{W}} |\phi_i - \phi^*_i|. \tag{3.22}$$

Lemma 13. *Given $Z^* \in \mathbf{L}$ and $Z \in \mathbf{C} \setminus \mathbf{L}$. Let Z' be a state not in the core which is reachable from Z by a least cost transition. Then there exists $Z_1 \in \mathbf{C}$ such that $D(Z^*, Z_1) < D(Z^*, Z)$ and $\mathbb{P}^t_0[Z', Z_1] > 0$ for some $t \geq 0$.*

Proof. By Theorem 12, the recurrent classes consist of all singleton states in $\mathbf{C}_0 \subseteq \mathbf{C}$. Thus it suffices to limit our analysis to $Z^* \in \mathbf{L} \cap \mathbf{C}_0$ and $Z \in \mathbf{C}_0 \setminus \mathbf{L}$ since other core states have zero-resistance paths to the states in $\mathbf{C}_0$.

Case A. Suppose that the least-cost transition to a non-core state is such that two (currently not matched) players experience trembles such that their match becomes profitable. That is, there exists i, matched to j, and a nonempty set J' such that i, j' is least costly to destabilize for any $j' \in J'$. Note that $d_i + d_{j'} - w_{ij'}$ is minimal for all $j' \in J'$ and thus constant and also the difference is non-negative since we are in a core state.

Case A.1. $d_i > d_i^*$.

Case A.1a. For all $j' \in J'$, $d_i + d_{j'} > w_{ij'}$.

We can construct a sequence of transitions such that i reduces his aspiration level by δ, j increases his aspiration level by δ (note that we have $d_j < d_j^*$), and all other aspiration levels stay the same. Note that D then decreased and the resulting state is again a core state given that for all $j' \neq j$ we started out with $d_i + d_{j'} > w_{ij'}$.

We shall explain the sequence in detail. Suppose the tremble occurs such that i reduces his aspiration level by at least δ and i and j' match at a price such that i's aspiration level does not increase. Consequently j and i' (j''s former partner, if he is matched in the core assignment) are now single. In the following period i and j are profitable. With positive probability, they match at a price such that d_i decreases by δ. Now i' and j' are both single. With positive probability they reduce their aspiration levels and rematch at their previous price, returning to their original aspiration levels. Thus, with positive probability the prices are set such that d_i decreases by δ, d_j increases by δ, and all other aspiration levels do not change. Hence D decreased and given the earlier observation the resulting state is again in the core, since now for all j', $d_i + d_{j'} \geq w_{ij'}$ and all other inequalities still hold.

For the subsequent cases we shall omit a description of the period by period transitions since they are conceptually similar.

Case A.1b. For all $j' \in J'$, $d_i + d_{j'} = w_{ij'}$.

It follows that $d_{j'} < d_{j'}^*$, hence a δ-reduction of i's aspiration level, a δ-increase of j's aspiration level, and δ-reductions and increases by all firms and workers on tight paths starting at i (in the same spirit as in Case 2a of the proof of Theorem 12) yield a reduction in D, and this leads to a core state.

Case A.2. $d_i = d_i^*$.

Since $Z \notin \mathbf{L}$ we must have $d_{j'} < d_{j'}^*$. For otherwise, given (i, j') is least costly to destabilize, we would have $Z \in \mathbf{L}$. But then j' must be matched in the core assignment and we have for j''s partner i' that $d_{i'} > d_{i'}^*$. Hence a δ-reduction in i' aspiration level and δ-increases by j' and all j'' for whom $d_{i'} + d_{j''} = w_{i'j''}$ yield a reduction in D, and this leads to a core state.

Case A.3. $d_i < d_i^*$.

We have $d_j > d_j^*$ and again a similar argument applies. A δ-reduction in j's aspiration level and δ-increases by i and all i' for whom $d_{i'} + d_j = w_{i'j}$ yields a reduction in D, and this leads to a core state.

Case B. Suppose that the least cost deviation to a non-core state is such that one player experiences a shock and therefore wishes to break up. That is, there exists i such that d_i is least costly to destabilize.

It follows that $d_i < d_i^*$, for otherwise $Z \in \mathbf{L}$ would constitute a contradiction.

Case B.1. For all $i' \neq i$, $d_{i'} + d_j > w_{i'j}$.

Again, we can construct a sequence of transitions such that i increases his aspiration level by δ, and j reduces his aspiration level by δ. Note that D then decreased and the resulting state is again in the core given that for all $i' \neq i$ we started out with $d_{i'} + d_j > w_{i'j}$.

Now we shall explain the sequence in detail. Suppose the shock occurs such that i turns single. Consequently j turns single too and, given that we are in a core state (i', j) is not profitable for any $i' \neq i$. Therefore, if j encounters any $i' \neq i$, he will reduce his aspiration level. Now i can rematch with his optimal match j at a new price such that i can increase his aspiration level by δ while d_j decreases his by δ. (Note that, for the latter transition, it is crucial that any matched couple has match value at least δ.) Hence D decreased and, given the earlier observation, the resulting state is again in the core, since now for all i' $d_{i'} + d_j \geq w_{i'j}$, and all other inequalities still hold.

Case B.2. There exists $I' \neq \varnothing$ and $i \notin I'$ such that for all $i' \in I'$, $d_{i'} + d_j = w_{i'j}$.

Similar to case A.1b a δ-increase of i's aspiration level, a δ-reduction of j's aspiration level, and δ-increases and reductions by all firms and workers on tight path starting at i (in the same spirit as in Case 2a of the proof of Theorem 12) yield a reduction in D, and this leads to a core state. ☺

Recall that the set of stochastically stable states is denoted by $\mathbf{S}$.

Theorem 14. *The stochastically stable states are maximally robust to one-period deviations, and hence* $\mathbf{S} \subseteq \mathbf{L}$.

Proof. We shall prove the theorem by contradiction. Suppose there exists $Z^* \in \mathbf{S} \setminus \mathbf{L}$. Let T^* be a minimal cost tree rooted at Z^* and suppose that $\rho(Z^*)$ is minimal. Let $Z^{**} \in \mathbf{L}$. By Lemma 13 together with the fact that the state space is finite, we can construct a finite path of least cost transitions between different core states such that their distance to a core state in $\mathbf{L}$ is decreasing:

$$Z^* \to Z_1 \to Z_2 \to \ldots \to Z_k = Z^{**} \tag{3.23}$$

Now we perform several operations on the tree T^* to construct a tree T^{**} for Z^{**}. First add the edges $Z_1 \to Z_2, \ldots, Z_{k-1} \to Z_k$ and remove the previously exiting edges from $Z_1, \ldots, Z_{k-1}$. Note that since the newly added edges are all minimal cost edges the sum of resistances does not increase. Next, let us add the edge $Z^* \to Z_1$ and delete the exiting edge from Z_k. Since $Z^* \notin \mathbf{L}$ it follows that $r(Z^* \to Z_1) < r(Z_k \to \cdot)$ and hence

$$\rho(Z^{**}) \leq \rho(Z^*) + r(Z^* \to Z_1) - r(Z_k \to \cdot) < \rho(Z^*) \tag{3.24}$$

This constitutes a contradiction. ☺

We can formulate natural conditions under which the stochastically stable set coincides with the least core:

Well-connected. An assignment game is *well-connected* if for any non-core state and for any player $i \in \mathcal{F} \times \mathcal{W}$ there exists a sequence of transitions in the unperturbed process such that i is single at its end.

Rich. An assignment game with match values $\boldsymbol{w}$ is *rich* if for every player $i \in \mathcal{F}$ there exists a player $j \in \mathcal{W}$ such that (i, j) is never profitable, that is, $w_{ij} = 0$.

Corollary 15. *Given a well-connected and rich assignment game with a unique core matching*[17]*, the set of stochastically stable states coincides with the least core, that is* $\mathbf{S} = \mathbf{L}$.

Proof. Given two recurrent classes of the process, $Z^*, Z^{**} \in \mathbf{C}_0$, and a non-core state, $Z \notin \mathbf{C}$, which is reachable from Z^* by a least cost deviation, we shall show that $r(Z, Z^{**}) = 0$. Suppose that Z^{**} has aspiration levels $\mathbf{d}^{**}$.

The idea of the proof is to construct a finite family of well-connected sequences, such that, after going through all transitions, players have aspiration levels less than or according to $\mathbf{d}^{**}$. Once we are in such a state Z^{**} can be reached easily.

Suppose we wish to make worker j_k single. By well-connectedness there exists a sequence of rematchings which makes j_k single at its end.[18] Suppose we have a minimal-length sequence among all sequences making j_k single, say $(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)$ if j_1 is matched and $j_1, (i_2, j_2), \ldots, (i_k, j_k)$ if j_1 is

[17] Generically the optimal matching is unique. In particular this holds if the weights of the edges are independent, continuous random variables. Then, with probability 1, the optimal matching is unique.

[18] Note that the sequence naturally needs to alternate between firms and workers in order to make players single along the way.

single. Note that it must hold that all players along the sequence (except potentially j_1) are currently matched for otherwise the sequence is not minimal. (This is the case since any single, by richness, reduces his aspiration level to zero and would then be a natural starting point for a shorter sequence as shall become clear below.) Further, by a similar observation, the sequence does not allow for profitable matches except for a match between j_1 and i_2. Finally, note that, for the sequence allowing j_k to become single at the end we must have that $d_{i_{l+1}} < \alpha_{i_{l+1}j_l}$ for $l = 1, \ldots, k-1$. This fact, together with the former comment, implies in particular that $d_{j_l} > 0$ for $l = 2, \ldots, k-1$.

Now suppose j_1 matches with i_2 (which is possible by assumption of well-connectedness) and j_2 becomes single. Suppose that the price is set such that d_{i_2} does not increase. Then, by richness, with positive probability j_2 reduces his aspiration level to 0. Suppose next that j_2 matches again with i_2 and the price is set such that j_2 keeps aspiration level 0. Now j_1 (and i_1 if j_1 was previously matched) is single and, by richness, with positive probability reduces his aspiration level to zero. Note that j_2 and i_3 are now profitable and they match with positive probability at a price such that d_{i_3} does not increase. Hence by iterating this sequence of transitions we arrive at a state where all players along the sequence are single except j_{k-1} and i_k who are now matched to each other. Suppose that they matched at a price such that j_{k-1} kept aspiration level 0. In particular note that j_k is single.

Next, we describe transitions such that the matching along the sequence considered remains the same (that is only (j_{k-1}, i_k) are matched) and the aspiration levels of j_{k-1} and i_k are such that $d_{j_{k-1}} \leq d^{**}_{j_{k-1}}$ and $d_{i_k} \leq d^{**}_{i_k}$. This can be achieved by matching i_{k-1} with j_{k-1} at a price such that j_{k-1} keeps aspiration level 0. Next, by richness, i_k reduces his aspiration level to zero. Finally j_{k-1} and i_k match again at a price such that the aspiration levels are less than or equal to their core aspiration levels in $\mathbf{d}^{**}$. (This can actually occur since otherwise the aspiration level vector $\mathbf{d}^{**}$ would not be pairwise stable, contradicting that Z^{**} is a core state.)

Now, by the well-connectedness assumption, we know that from any non-core state and for every player there exists such a sequence. Hence successively applying sequence after sequence such that each player is at its end once, we can conclude that there exists a path to a state such that, for all i, $d_i \leq d_i^{**}$. But note that some players may be matched.

Next, we have to show how Z^{**} is reached from the latter state. This is achieve if we successively match all (i,j) who are matched in Z^{**}, who are not matched yet, and for whom $d_i + d_j < \alpha_{ij}$ (they are profitable) at a price such that their new aspiration levels are d_i^{**}, d_j^{**}. (Here we need the assumption that the core matching is unique. Given our construction, players may be matched at the end of the sequences described above. In particular the core might already be reached with aspiration levels d^{**}. Now, if the optimal matching is not unique, we cannot guarantee that any particular matching is in place.) This leads to a state where aspiration levels are $\mathbf{d}^{**}$. Note that these aspiration levels are good. Further note that a reduction of the sum of aspiration levels will lead to a state which is not good. Cases 1a,b and 2b of the proof of Claim 2 of Theorem 12 can now be applied iteratively (Case 2a cannot hold, otherwise aspiration levels will no longer be good). These cases change matchings but do not change aspiration levels. Hence eventually the desired core state Z^{**} is reached.

The proof summarizes as follows. We have shown that once the process is in a non-core state any core state can be reached. Hence the analysis of stochastic stability reduces to the resistance of exiting a core state. But this resistance is uniquely maximized by the states in the least core, which thus coincides with the set of stochastically stable states. ☺

3.6.2 Example

We shall illustrate the predictive power of our result for the 3×3 'house trade game' studied by Shapley and Shubik (1972). Let three sellers (w_1, w_2, w_3) and three buyers (f_1, f_2, f_3) trade houses. Their valuations are as follows:

Table 3.1: Seller and buyer evaluations.

House	Sellers willingness to accept	Buyers' willingness to pay		
j	$r_j^+(1) = r_j^+(2) = r_j^+(3)$	$r_1^+(j)$	$r_2^+(j)$	$r_3^+(j)$
1	$18,000$	$23,000$	$26,000$	$20,000$
2	$15,000$	$22,000$	$24,000$	$21,000$
3	$19,000$	$21,000$	$22,000$	$17,000$

These prices lead to the following match values, w_{ij} (units of 1,000), where sellers are occupying rows and buyers columns:

$$W = \begin{pmatrix} 5 & \mathbf{8} & 2 \\ 7 & 9 & \mathbf{6} \\ \mathbf{2} & 3 & 0 \end{pmatrix} \tag{3.25}$$

The unique optimal matching is shown in bold numbers. Shapley and Shubik (1972) note that it suffices to consider the 3-dimensional imputation space spanned by the equations $d_{w_1} + d_{f_2} = 8$, $d_{w_2} + d_{f_3} = 6$, $d_{w_3} + d_{f_1} = 2$. Figure 3.8 shows the possible core allocations.

We shall now consider the least core, $\mathbf{L}$. Note that the particular states in $\mathbf{L}$ depend on the step size δ. We shall consider $\delta \to 0$ to best illustrate the core selection. By an easy calculation one finds that the states which are least

Figure 3.8: Imputation space for the sellers.

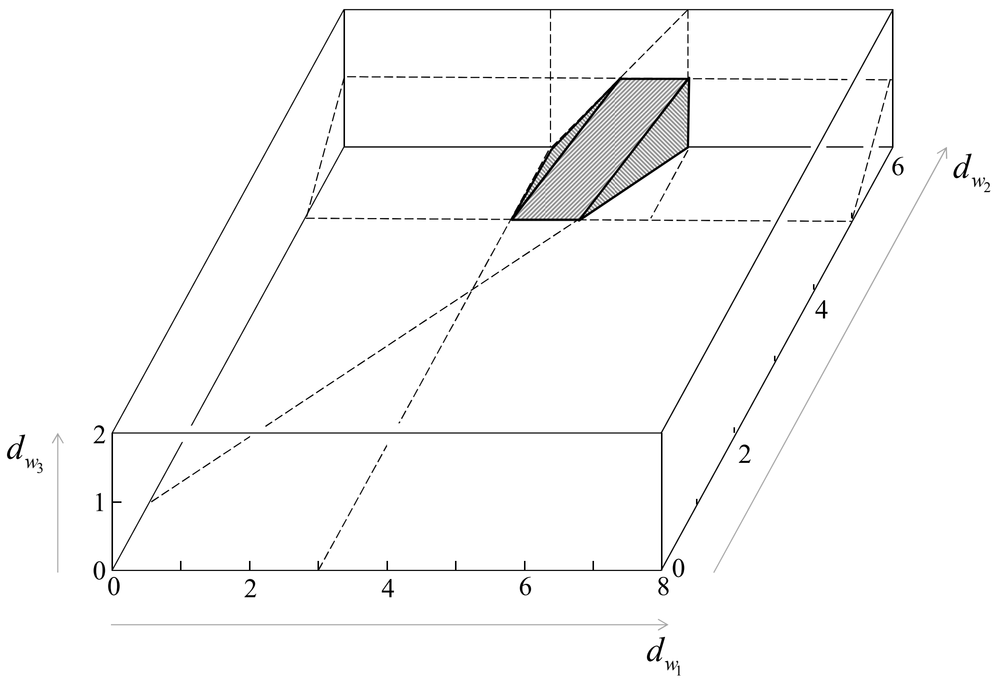
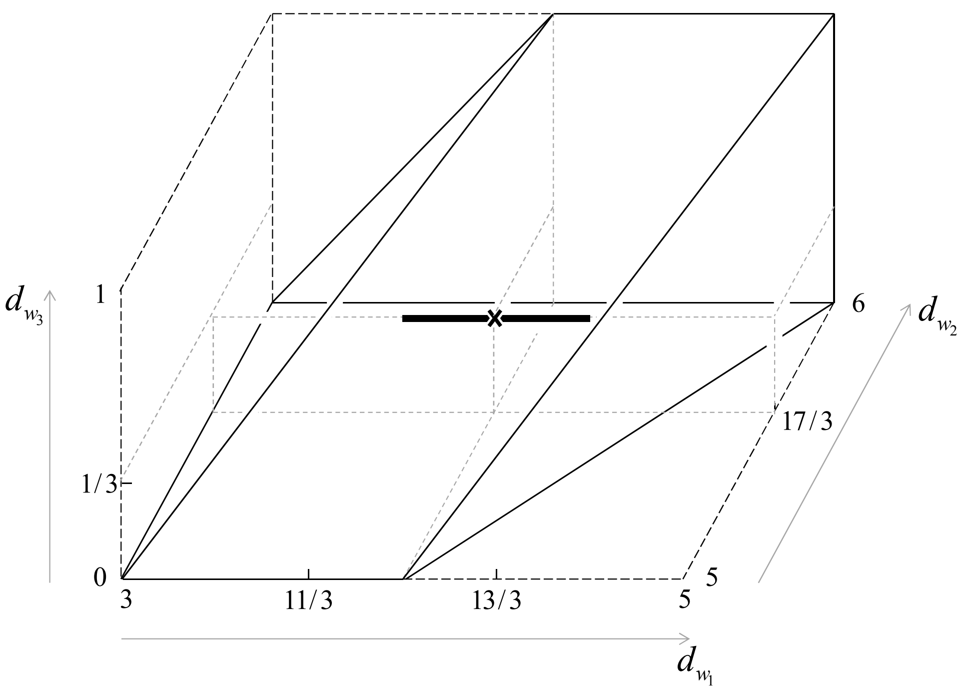


Figure 3.9: Core selection for the sellers.



vulnerable to one-period deviations are such that

$$d_{w_1} \in [11/3, 13/3],\ d_{w_2} = 17/3,\ d_{w_3} = 1/3. \tag{3.26}$$

The minimal excess in the least core is $e_{\min} = 1/3$. The bold line in Figure 3.9 shows the set $\mathbf{L}$. The nucleolus, $d_{w_1} = 4$, $d_{w_2} = 17/3$, $d_{w_3} = 1/3$, is indicated by a cross. (One can verify, that here the kernel coincides with the nucleolus.)

3.7 Conclusion

In this chapter we have shown that agents in large decentralized matching markets can learn to play equitable core outcomes through simple trial-and-error learning rules. We assume that agents have no information about the distribution of others' preferences, their past actions and payoffs, or the value of different matches. The unperturbed process leads to the core with probability one but no authority 'solves' an optimization problem. Rather, a path into the core is discovered in finite time by a random sequence of adjustments by the agents themselves. This result is similar in spirit to that of Chen et al. (2011), but in addition our process selects equitable outcomes within the core. In particular, the stochastically stable states of the perturbed process are contained in the least core, a subset of the core that generalizes the nucleolus for assignment games. This result complements the stochastic stability analysis of Newton and Sawa (2015) in ordinal matching and of Newton (2012) in more general coalitional games. It is an open problem to extend the analysis to more general classes of cooperative games and matching markets.

Chapter 4

Decentralized dynamics and fast convergence in assignment games

4.1 Introduction

In the previous chapter we showed that there exist learning rules which lead to the core despite the following restrictions:

Market decentralization. Agents interact, match, and break up at random points in time. That is, the order of interactions is random and further there is no central price setting rule.

Information decentralization. Players have no information about the value other players attach to being matched with them. Moreover they cannot deduce such information from prior rounds of play. There is no central authority with such knowledge either.

As discussed in section 3.1 many markets exhibit these low levels of information together with a high degree of decentralization. On the other hand some markets (such as labor markets) are influenced by macro-economic signals. Therefore we propose that players are able to observe, along with the current market prices, a general market sentiment, for example whether we are in times of crisis or boom.

In this chapter we propose an adaptive process that reflects the decentralization and thus the limited information agents have about the market. We show that this dynamic leads *in polynomial time* to the core of the assignment game (Shapley and Shubik 1972). Polynomial time convergence is an important feature for behavioral dynamics to have explanatory bite for large markets. In addition it is crucial from a market design perspective. From a technical perspective, we develop new results for random walks on graphs which may be of use more generally. In particular, to our knowledge, we are the first ones to use random walk arguments in the context of assignment and matching markets. Note that in the previous chapter we did not attempt to answer the question of fast convergence.

The next section discusses the related literature some of which has already been mentioned in section 3.2. In section 4.3 we recall the underlying model for bilateral markets with transferable utility and introduce our dynamic model. We also provide an example. Section 4.4 states our main results and a discussion of its assumptions. Section 4.5 first provides an intuition and then states the formal proof. The proof develops novel results for random walks on graphs. Section 4.6 concludes.

4.2 Related literature

Matching theory is a flourishing area of study in cooperative game theory and mechanism design.[1] A prominent example of a cooperative game is the *assign-*

[1] See, for example, Roth and Sotomayor (1990) and Peleg and Sudholter (2007) for introductions.

ment game where two sides of the market, firms and workers, form bilateral coalitions and share the 'cooperative value' according to some sharing rule. Shapley and Shubik (1972) show that there always exists an assignment such that the total sum of utilities is maximized and the division of surplus is robust to coalitional deviations. The set of states with these properties is called the *core*. One of the first algorithms to find a solution in the core of the assignment game has time complexity of $O(N^4)$, where N is the number of players (Kuhn 1955). Edmonds and Karp (1972) give a tighter bound of $O(N^3)$ which, to our knowledge, is the best-known bound for a centralized algorithm.[2]

Our results fit into a growing literature showing how cooperative game solutions can be reached by non-cooperative dynamic processes (Agastya 1997, 1999, Arnold and Schwalbe 2002, Rozen 2013, Newton 2010, 2012, Sawa 2014, Newton and Sawa 2015). Several recent studies consider the convergence for a family of *decentralized* dynamics for the assignment game that satisfy our definitions outlined in section 4.1. Chen et al. (2011), Nax and Pradelski (2014), Biró et al. (2012), Klaus and Payot (2013), and Nax et al. (2013b) show that specific decentralized dynamics converge to the core in finite time with probability one from any initial state.[3] However, none of these papers establishes polynomial time convergence which is the subject of this chapter.

There is another strand of the literature that establishes polynomial time convergence, but the processes are not decentralized in our sense. *Belief propagation* is a message passing algorithm for computing optimal matchings and allocations on graphs.[4] In each time step all agents send messages to each other indicating how much they are willing to pay at most in order to be matched

[2]Seminal subsequent papers include Crawford and Knoer (1981), Kelso and Crawford (1982), Demange and Gale (1985), Demange et al. (1986).

[3]Note that Biró et al. (2012) holds for one-to-one matching on general graphs. It thus includes the assignment game (one-to-one matching on bipartite graphs) as a special case.

[4]See, for example, Pearl (1982), Kim and Pearl (1983), Pearl (1988), Richardson and Urbanke (2001).

with one another. Bayati et al. (2005, 2008) and Salez and Shah (2009) show that, for bipartite graphs with a unique optimal matching, their algorithms converge in polynomial time to a stable allocation. The matching is thus enforced by the allocation, but it is not specified how players know when they should match. In the same spirit *bargaining dynamics* have been studied (see Kleinberg and Tardos 2008). In addition to sending messages about the maximal amount a player is willing to pay, there is a fixed sharing rule (50 – 50 above outside-options). Celis et al. (2010) and Bayati et al. (2014) show that, for one-to-one matching markets,[5] their algorithms converge in polynomial time to a subset of the core which approximates the generalized Nash-bargaining solution (Rochford 1984). Note that these models are not decentralized in our sense: First, for *market decentralization*, all agents play simultaneously according to a deterministic rule. In particular there is a central price setting mechanism and a clearing house. Second, for *information decentralization*, the updating rule requires agents to know the valuations other players hold for them. Interestingly, Azar et al. (2009) show that the afore mentioned bargaining dynamics takes exponentially long to converge if market decentralization holds, that is, if players send messages at random points in time.

There is also a related literature on decentralized dynamics for the marriage problem (Gale and Shapley 1962, see Roth and Sotomayor 1990 for a survey). The key difference between the marriage problem and the assignment game is that the former studies a matching market with *non-transferable utility*: There is no medium of exchange and players have ordinal utility functions. On the other hand, the assignment game is a model of *transferable utility*: There exists a medium of exchange (such as money) in which players' cardinal utility functions are linear. For the marriage problem there exist decentralized adjustment processes leading to stable outcomes in finite time (Roth and

[5] This result holds for one-to-one matching on general graphs.

Vande Vate 1990).[6] Ackermann et al. (2011) recently showed by example that the dynamic by Roth and Vande Vate (1990) may take exponentially long to converge.

The contribution of the present chapter is to demonstrate that a simple decentralized adjustment process leads to the core of the assignment game in polynomial time. Our proof develops novel results for random walks on graphs, and more generally suggests a fruitful connection between the theory of random walks and matching theory.

4.3 Model

Our model is in the spirit of chapter 3 (which we refer the reader to for its behavioral motivation): Matched agents occasionally try to re-match in order to increase their payoff, while single agents, in the hope of attracting partners, lower their aspiration levels and are willing to match at a lower payoff.

In the next subsection we shall introduce the well-known assignment game (Shapley and Shubik 1972) and the dynamic components of our model. Subsection 4.3.2 explains the dynamics of play and how step-by-step transitions work. In subsection 4.3.3 we give an example.

4.3.1 The assignment game

Firms $i \in \mathcal{F}$ and workers $j \in \mathcal{W}$ interact by matching, breaking up, and rematching in the hope of achieving higher payoffs. Let $V = \mathcal{F} \cup \mathcal{W}$. We assume that $|\mathcal{F}| = |\mathcal{W}| = N$. (Note that this can always be achieved by adding 'dummy'

[6]See Chung (2000), Diamantoudi et al. (2004), Kojima and Ünver (2008), Klaus and Klijn (2007) for subsequent results on more general families of games.

agents to the smaller side of the market.) A player can be *matched* to at most one player at a given point in time. A player currently not matched to another player is called *single*.

Reservation value. Each firm i has a *reservation value*, $r_i^+(j)$ for being matched with worker j. That is, firm i is willing to pay at most $r_i^+(j)$ when matching with worker j and otherwise prefers to remain single. Similarly each worker j is willing to accept no less than $r_j^-(i)$ when matching with firm i and otherwise prefers to remain single.

Match value. The *match value* of a match (i,j) is

$$w_{ij} = (r_i^+(j) - r_j^-(i))_+. \tag{4.1}$$

Let $W = (w_{ij})_{i\in\mathcal{F},j\in\mathcal{W}}$. Let $w^* = \max_{i,j} w_{ij}$.

Note that, if *information decentralization* holds, the reservation values are private information and not known to other agents. Then, w_{ij} is a hidden variable and is not known to any of the players including i and j.

Since we will be interested in dynamic processes we introduce time steps $t = 0, 1, 2, 3, \ldots.$

Matching. The *matching* at time t is described by $M^t = (m_{ij}^t)_{i\in\mathcal{F},j\in\mathcal{W}} \in \{0,1\}^{\mathcal{F}\times\mathcal{W}}$ such that each row and each column of M^t sums to at most 1.

Price. When i is matched with j at time t they trade at a unique *price*, π_{ij}^t.

Payoff. If (i,j) are matched with each other the *payoff* to firm i is $\phi_i^t = r_i^+(j) - \pi_{ij}^t$ and the payoff to worker j is $\phi_j^t = \pi_{ij}^t - r_j^-(i)$. If player i is single his *payoff* is $\phi_i^t = 0$. Let $\Phi^t = (\phi_i^t)_{i\in V}$. Note that the payoffs are private information.

The *matching market* is described by $[V = \mathcal{F} \cup \mathcal{W}, W, M^t]$. Then, given the matching market, the *assignment game* $G(\nu, V)$ is defined as in chapter 3.3.1.

We shall now define desirable properties of outcomes:

Optimality. A matching M is *optimal* if $\sum_{i,j \in V} m_{ij} \cdot w_{ij}$ is maximal.

Pairwise stability. A payoff vector Φ is *pairwise stable* if for all $i, j \in V$: $\phi_i + \phi_j \geq w_{ij}$.

Core. A matching M together with the payoffs Φ is in the *core* if M is optimal and Φ is pairwise stable.

So far we introduced the assignment game without any additional components. We shall now introduce dynamic components of our model.

Aspiration levels. At the end of any step $t = 0, 1, 2, \ldots$, each player $i \in V$ has an *aspiration level*, $d_i^t \geq 0$, which determines the minimal payoff at which he is willing to be matched. Let $\mathbf{d}^t = (d_i^t)_{i \in V}$.

Feasibility. A pair of aspiration levels (d_i^t, d_j^t) is *feasible* if, there exists a price such that if i and j match, at least one of the two players has a higher payoff. Formally, if $r_i^+(j) - d_i^t > r_j^-(i) + d_j^t$ (equivalently $d_i^t + d_j^t < w_{ij}$) or if $r_i^+(j) - d_i^t = r_j^-(i) + d_j^t$ (equivalently $d_i^t + d_j^t = w_{ij}$) and one of the players is single.[7] We shall say that i and j are *feasible* in period t if (d_i^t, d_j^t) is feasible.

Market sentiment and price flexibility. *Market sentiment* is an exogenous binary random variable, $\Psi^t \in \{\ominus, \oplus\}$, that acts as a correlation device determining which players currently exhibit *price flexibility.* If market sentiment is negative ($\Psi^t = \ominus$), for example in periods of high unemployment, workers

[7]Note that this definition assumes a tie-breaking rule: A player prefers to be matched over being single and matches are broken up in favor of ones of equal value and with one unmatched player.

exhibit *price flexibility.* If market sentiment is positive ($\Psi^t = \oplus$), for example in periods of low unemployment, firms exhibit *price flexibility.*

No better alternative. Suppose (i, j) are feasible. Then say that player j has *no better alternative than* i if for all $i' \neq i$ $d^t_{i'} + d^t_j > w_{i'j}$. Otherwise say that j *has a better alternative than* i.

An agent who is not price flexible only agrees to a match if the price is such that his payoff will be at least as high as his aspiration level. On the other hand, an agent j who is currently price flexible will be less tough in negotiations with agent i if j has no better alternative than i. In this case j may agree to a price such that his payoff is a bit below his aspiration level.

State. The state at the end of period t is given by the matching, the aspiration levels, and market sentiment: $[M^t, \mathbf{d}^t, \Psi^t]$.

Finally, assume that there is a minimal unit of payoff $\delta \in \mathbb{R}^+$, and all match values, aspiration levels, etc. are multiples of δ.

4.3.2 Dynamic play

Agents are *activated* by independent, identical Poisson arrival processes. Once a player is activated the next discrete time period $t + 1$ starts:

The active agent scans the other side of the market, looking for a feasible match. For example, consider a job-list available online. If there is a feasible match he picks one at random and matches with him. The price is set at random satisfying the conditions determined by the aspiration levels and price flexibility. We can think of the price as the outcome of a bargaining process that we do not model explicitly. At the end of each period the two players adapt their aspiration levels to equal their new payoffs. If there is no feasible match and the active player is matched the state remains the same. If there is no feasible match and the active player is single he remains single and reduces his aspiration level.

Formally, let $[M^0, \mathbf{d}^0, \Psi^0]$ be any initial state. Suppose that in period $t+1$ ($t \geq 0$) firm i is activated (the dynamic is symmetric if a worker is activated).

If there exists a feasible match for i, i picks one such player at random, matches with him and the price π_{ij}^{t+1} is set at random with full support between

$$r_j^-(i) + d_j^t - \Delta_j(i) \quad \text{and} \quad r_i^+(j) - d_i^t, \tag{4.2}$$

where

$$\Delta_j(i) = \begin{cases} \delta & \text{if } j \text{ is price flexible and has no better alternative,} \\ 0 & \text{else.} \end{cases} \tag{4.3}$$

At the end of the period the aspiration levels of the newly matched pair (i, j) are set to their new realized payoffs,

$$d_i^{t+1} = \phi_i^{t+1} = r_i^+(j) - \pi_{ij}^{t+1} \qquad \text{and} \qquad d_j^{t+1} = \phi_j^{t+1} = \pi_{ij}^{t+1} - r_j^-(i), \tag{4.4}$$

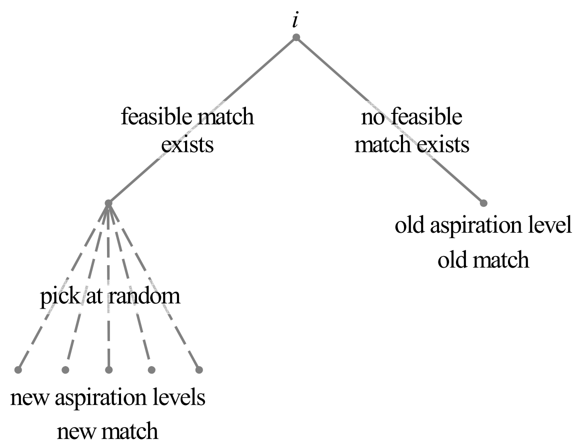
and the matching M^{t+1} is updated accordingly, that is, $m_{ij}^{t+1} = 1$, $m_{kl}^{t+1} = 0$ for all $k \neq i$ with $l = j$ and for all $l \neq j$ with $k = i$.

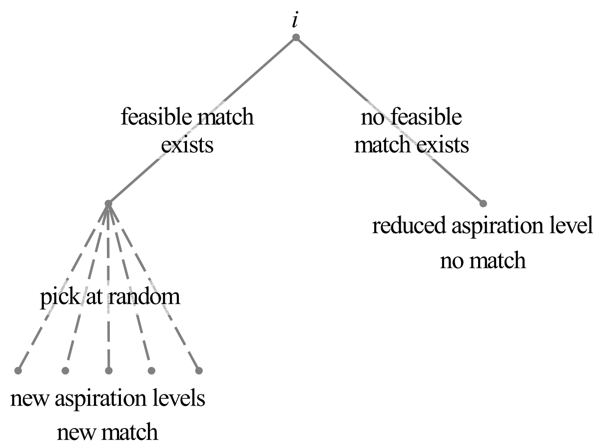
If there is no feasible match for player i, if i was previously matched he stays with his previous partner and keeps his payoff and thus aspiration level. Otherwise i remains single and reduces his aspiration level:

$$d_i^{t+1} = d_i^t - \delta \tag{4.5}$$

See Figure 4.1 for illustrations.

Figure 4.1: Transition diagram for active, matched (top) and active, single agent (bottom).



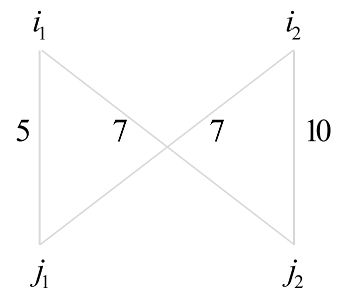


Note that, *market and information decentralization* hold. For market decentralization, agents interact at random points in time and prices are set at random with weak boundary conditions which are natural from the point of view of each player. For information decentralization, in each step players only know their own reservation values, aspiration level, and the price they trade at. They do not know other players' reservation values or aspiration levels.

Finally, given the updating rule, the payoffs and prices are deducible from the aspiration levels and the matching. For ease of exposition we shall focus on the latter.

4.3.3 Example

Let $V = \mathcal{F} \cup \mathcal{W} = \{i_1, i_2\} \cup \{j_1, j_2\}$ and $w_{11} = 5, w_{12} = w_{21} = 7, w_{22} = 10$. Let $\delta = 1$.

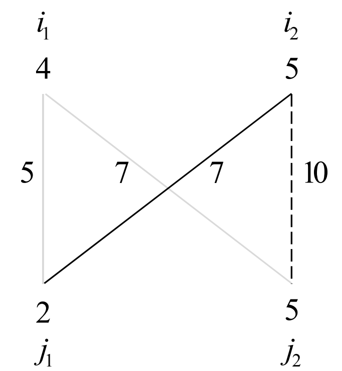


In all figures a gray edge between two players i and j indicates that $w_{ij} > 0$, a black solid edge that they are matched ($m^t_{ij} = 1$), and a black dashed edge that $d^t_i + d^t_j = w^t_{ij}$ and i and j are not matched in period t. We shall call such a dashed edge *tight*. Numbers next to edges indicate match values. The current aspiration levels are shown next to the name of an agent.

One can verify that the optimal matching is unique and such that firm i_1 matches worker j_1 and i_2 matches j_2.

period t: *Current state*

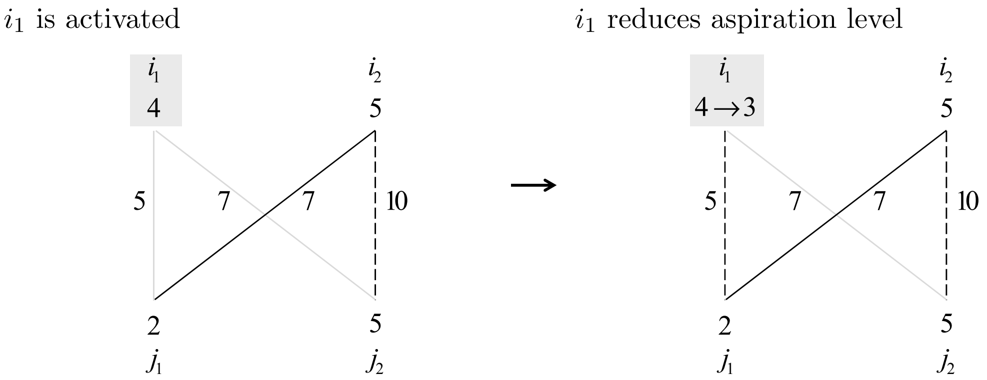
Suppose that, at the end of some period t, (i_2, j_1) are matched, i_1 and j_2 are single, and i_2 and j_2 share a tight edge.



Note that the aspiration levels satisfy $d_i^t + d_j^t \geq w_{ij}$ for all i and j, but the assignment is not optimal and the aspiration levels (and thus payoffs) are not stable.

period $t+1$**:** *Activation of single agent* i_1 *and negative market sentiment* $\Psi^t = \ominus$

Suppose that i_1 is activated in period $t+1$ and market sentiment is currently negative. i_1's current aspiration level is too high in order to be feasible with any worker. Hence he remains single and reduces his aspiration level by 1.

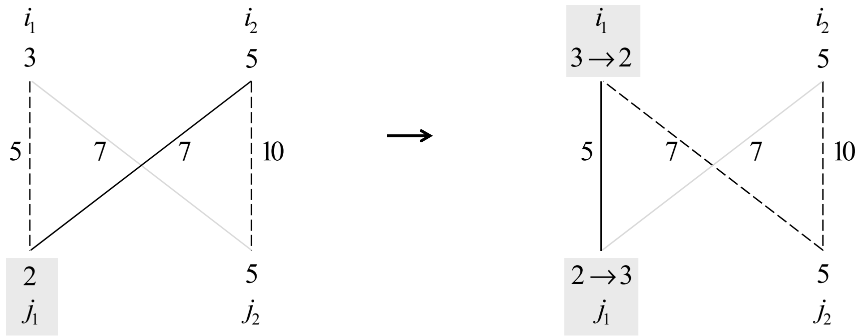


period $t+2$**:** *Activation of matched agent* j_1 *and positive market sentiment* $\Psi^{t+1} = \oplus$

Suppose that j_1 is activated in period $t+2$ and market sentiment has changed and it is now positive. (i_1, j_1) is the only feasible pair involving j_1. They thus match and given that i_1 has no better alternative and is price flexible (since market sentiment is currently positive) the price, with positive probability, is set such that j_1 increases his aspiration level by $\delta = 1$.

j_1 is activated

Successful match; j_1 increases aspiration level, i_1 reduces

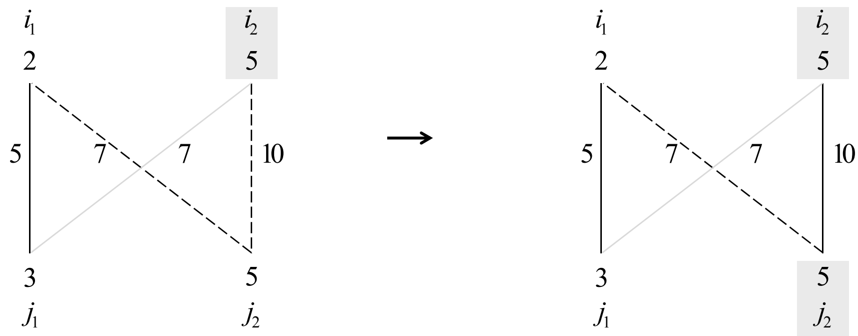


period $t+3$**:** *Activation of single agent* i_2 *and negative market sentiment* $\Psi^{t+2} = \ominus$

Suppose that i_2 is activated in period $t+3$ and market sentiment has changed again is now negative. (i_2, j_2) is the only feasible pair involving i_2. They thus match and although workers are currently price flexible (market sentiment is negative), j_2 has a better alternative (namely i_1) and thus is not accepting less than his aspiration level. They are matching at their current aspiration levels.

i_2 is activated

Successful match with j_2 at current aspiration levels



Note that the resulting state is in the core. We invite the reader to convince himself that this state is absorbing.

4.4 Results

Since the model lives on the Cartesian product of the discrete grid $\delta \cdot \mathbb{Z}$, the term *generic* needs to be treated with care. We shall give an intuitive definition which will be sufficient for our results to hold.

Definition 16. *The match values W are* discrete generic *if there is no sequence of firms and workers $(1, 1', 2, 2', ..., k, k')$, $\{1, \ldots, k\} \subseteq \mathcal{F}, \{1', \ldots, k'\} \subseteq \mathcal{W}$ such that*

$$w_{11'} + w_{22'} + ... + w_{kk'} = w_{1k'} + w_{21'} + ... + w_{k(k-1)'}. \tag{4.6}$$

That is, the corresponding graph has no cycles such that there are two alternating partitions which have the same sum of match values.[8]

Theorem 17. *Given a discrete generic assignment game, suppose that Ψ^t switches at a rate of $\Theta(N^{2+c})$ $(c \geq 0)$ time steps.*[9] *In expectation, the dynamic converges to the core in $O(N^{3+c} \cdot \max\{N, \frac{w^*}{\delta}\})$ iterations with total computational complexity of $O(N^{5+c} \cdot \max\{N, \frac{w^*}{\delta}\})$.*

Note that for $c = 0$ the rate of switching for Ψ^t provides a lower bound, $c > 0$ will slow down the process but the dynamic still converges in polynomial time. On the other hand, if both sides of the market are always price flexible or market sentiment switches too fast the rate of convergence is exponential:

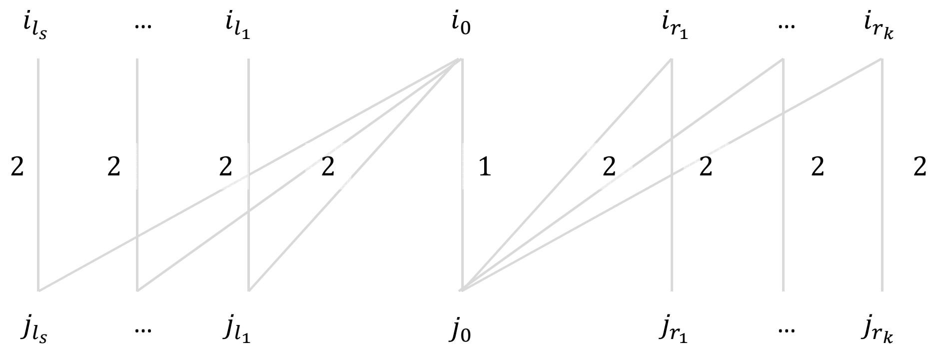
Proposition 18. *Suppose that Ψ^t switches at a faster rate than $O(N^2)$ time steps or both sides of the market always exhibit price flexibility. There exists a discrete generic assignment game such that the expected rate of convergence is $O(e^N)$.*

[8] Note that, given we consider bipartite graphs, there are no odd cycles.

[9] $f(N) = \Theta(g(N))$ if there exist two constants $k, K \geq 0$ and a positive integer N_0 such that $kg(N) \leq f(N) \leq Kg(N)$ for all $N \geq N_0$.

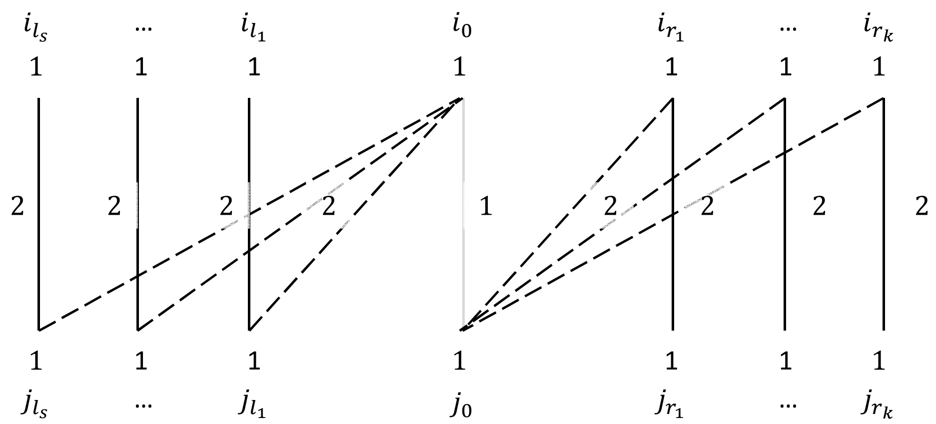
Proof. We shall prove this by constructing an example. Suppose we have an assignment game as shown in Figure 4.3 and the indexes s, k are both $O(N)$. The unique optimal matching is given by the vertical lines. Suppose that the minimal unit is $\delta = 1$.

Figure 4.3: Counterexample – necessity of alternating market sentiment (1).



Suppose we are in the state displayed in Figure 4.4, where the aspiration levels are shown next to each agent. Note that the only edge such that the sum of the aspiration levels exceeds the match value is (i_0, j_0).

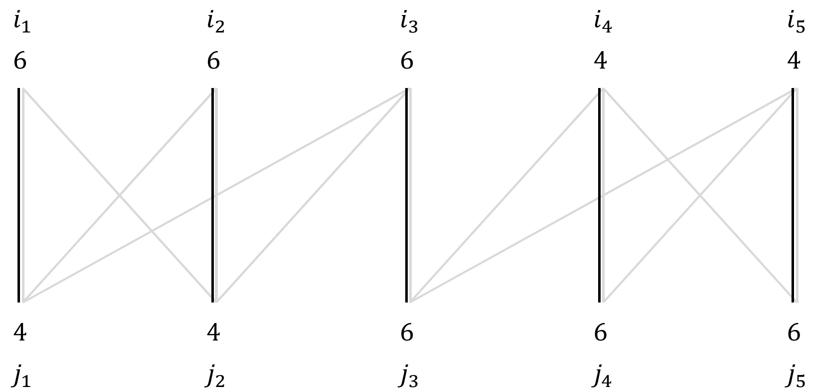
Figure 4.4: Counterexample – necessity of alternating market sentiment (2).



In order for i_0 and j_0 to match at least one of them has to reduce his aspiration level. But this will only be the case if they have no better alternative. Thus either all of $i_{r_1}, \ldots, i_{r_k}$ or all of $j_{l_1}, \ldots, j_{l_s}$ have to increase their aspiration levels simultaneously. If the agents on both sides of the market are price flexible the sharing between any two (currently matched) agents switches back and forth between sharing equally $(1 - 1)$ and sharing $0 - 2$ with the player previously having a tight edge with i_0 or j_0 receiving 2. This assumes that the players $i_{r_1}, \ldots, i_{r_s}$ and $j_{l_1}, \ldots, j_{l_k}$ have a new tight edge to some player (not displayed for ease of exposition) once they reduced their aspiration level to zero. The probability of them all being at aspiration level 2 at the same time is of the order of $\frac{1}{2^N}$ and therefore the waiting time is exponential and is $O(2^N)$. ☺

Next, we shall argue that it is important to restrict the games under consideration by a genericity assumption. Consider the example in Figure 4.5 which is not discrete generic. Suppose all gray edges have match value 10 and no other edges exist. Let $\delta = 1$. The numbers below/above the players give their current aspiration level. Note that each player, say i, facing another player, j, has a better alternative and for all i, j $d_i + d_j \geq w_{ij}$ holds. Thus regardless of the current matching, aspiration levels will not change in this state. This is the case since a player only increases his aspiration level if he receives a higher payoff. But given that for all i, j $d_i + d_j \geq w_{ij}$ a payoff increase can only occur if another player suffers a decrease in his payoff and thus aspiration level. But in light of the price rule no player reduces his aspiration level facing another player as long as he has a better alternative. The unique core state is such that players i_3 and j_3 are matched (indicated by black dotted lines in the figure) but their aspiration levels do not permit them to be matched. Therefore the process does not reach the core.

Figure 4.5: Counterexample – Weight matrix is not discrete generic.



4.5 Proof of the main result

We shall prove Theorem 17 via several intermediate statements. We first give the intuition and then the formal statements and proofs.

Note that the process is fully characterized by the state variables M^t, d^t, and Ψ^t. d^t together with the matching imply the payoffs and prices. We shall thus only focus on the changes in the state variables. We introduce notation in the spirit of chapter 3.5. An edge between player i and j is called *tight* if $d_i^t + d_j^t = w_{ij}$ and *loose* if $d_i^t + d_j^t < w_{ij}$. Given aspiration levels $\mathbf{d}^t$, if for all $i, j, d_i^t + d_j^t \geq w_{ij}$ and for each i, either there exists j such that $d_i^t + d_j^t = w_{ij}$ or $d_i^t = 0$ we shall call $\mathbf{d}^t$ *good*. Let a *tight component* be a maximal subgraph of players connected by tight edges.

4.5.1 Proof outline

First, we show that once a state is reached such that no loose edges exist all future states won't have loose edges (Lemma 19). We give a bound for the waiting time of $O(N^2 \log N)$. This crude bound results by noting that if two players meet once, they will never again share a loose edge. Thus it suffices that all pairs meet at least once.

Next, given we have reached the set of states with no loose edges the sum of aspiration levels is non-increasing in all future states (Lemma 20). We go on to bound the time until the sum of aspiration levels is minimal provided that there are no loose edges (Lemmas 21 and 22). We show that the minimum will be equal to the value of the game, $\nu(V)$. The proof of this result makes use of market sentiment and the assumption that the game is discrete generic (see Definition 16). By discrete genericity all tight components are trees. If all tight components allow perfect matchings we are done. Otherwise we show how in a given tight component which does not allow a perfect matching the sum of aspiration levels can be reduced. We use the market condition to prove that the expected time until the sum of aspiration levels is equal to the value of the game is $O(N^3 \cdot \frac{w^*}{\delta})$. Note that once we are in such a state we will always remain in the set of states with no loose edges and the sum of aspiration levels equal to the value of the game (by Lemmas 19 and 20).

Next, we show by contradiction that from now on for all optimal pairs we have that the edge between them is tight (Lemma 23). It follows that the current aspiration levels are already corresponding to a core allocation. Thus only matchings, break-ups and rematchings are necessary in order to reach the core.

Random walk argument. In order to give a bound on the expected time until matchings occur we create a duality with random walks on trees (Lemma 24). We 'give' each single player a *single token* which is moved among the players as play unfolds. The idea is that the random walk governing the single tokens will behave in the same way as the underlying model. That is if a player has a single token he is indeed single. If a single player i matches a player from the other side of the market j, the single token is moved to j's previous match. We prove a bound for the meeting time of two random walks on (different) trees. By carefully coupling the movements in our model with the two random walks we show that the expected time until the core is reached from a state as discussed above is $O(N^4)$. We then show that any core state is absorbing.

Finally, it is easy to see that a single iteration has computational complexity $O(N^2)$ (Lemma 25). Putting together the different bounds Theorem 17 will follow.

4.5.2 Formal proof

Lemma 19. *Starting from any initial state $[M^0, \mathbf{d}^0, \Psi^0]$, the expected time until the dynamic reaches a state such that for all i, j $d_i + d_j \geq w_{ij}$ is $O(N^2 \log N)$. Once the latter property holds for some $T \geq 0$ it holds for all $t \geq T$.*

Lemma 20. *Let $D^t := \sum_{i \in V} d_i^t$. Suppose the process is in a state as in Lemma 19 at time $T \geq 0$. Then D^t is non-increasing for all $t \geq T$.*

Lemma 21. *Given a state as in Lemma 19, either the expected waiting time until D^t decreases by δ is bounded above by $O(N^2)$ or $D^t = \nu(V)$.*

Lemma 22. *Given a state as in Lemma 19, the expected waiting time until $D^t = \nu(V)$ is $O(N^3 \cdot \frac{w^*}{\delta})$.*

Lemma 23. *Given a state $[M^T, \mathbf{d}^T, \Psi^T]$ as in Lemma 22 let M be the optimal matching (which is unique by genericity). Then for all $(i, j) \in M$ and for all $t \geq T$, $d_i^t + d_j^t = w_{ij}$.*

Lemma 24. *Given a state as in Lemma 22, the expected time until the core is reached is $O(N^4)$. Further, any core state is absorbing.*

Lemma 25. *One iteration of the dynamic takes $O(N^2)$ operations.*

Proof of Lemma 19. It follows from the updating rule that for any $i \in \mathcal{F}, j \in \mathcal{W}$ and for some T, $d_i^T + d_j^T \geq w_{ij}$ holds, then for all $t \geq T$, $d_i^t + d_j^t \geq w_{ij}$, proving the second part of the Lemma. Now we use an analogy to the Coupon Collector's Problem (see, for example, Feller 1950) to bound the expected time until each pair of players have encountered each other. The Coupon Collector's Problem asks the following question: Suppose you have x coupons, what is the expected time until you have picked each coupon at least once when picking with replacement. The answer is that the number of trials needed is $O(x \log x)$. We thus have that the expected time until each pair of players (N^2) has been selected at least once is $O(N^2 \log N)$. ☺

Proof of Lemma 20. Given D^t, D^{t+1} can only increase if two players, say i, j, newly match and $d_i^t + d_j^t < w_{ij}$. But given we are in a state as in Lemma 19, for all i, j $d_i^t + d_j^t \geq w_{ij}$ and hence we conclude that $D^{t+1} \leq D^t$. ☺

Next we shall recall a folk result regarding the existence of perfect matchings on trees (see, for example, Edmonds 1965).

Lemma 26. *A tree C admits a perfect matching (a matching such that each player is matched) if and only if for every vertex v, the graph $C - v$ has exactly one odd sub-component.*

Proof. First, suppose there exists a perfect matching in C. Let u be the vertex matched with v in the perfect matching. Then each component in $C - v$ not containing u must have a perfect matching and is thus even. Finally the component containing u must be odd.

Second, suppose that for every vertex v the graph $C - v$ has exactly one odd component. We shall construct a perfect matching. Given C is a tree and by our assumption each vertex v has a unique neighbor in one component of odd size, say u. We shall show that pairing each v to its neighbor u in the odd component yields a perfect matching. It suffices to show that if u is the neighbor of v in the unique odd component of $C - v$ (say $odd(C - v)$) that it is also the case that v is the neighbor of u in the unique odd component of $C - u$ (say $odd(C - u)$). Note that the component of $C - u$ which contains v is disjoint with $odd(C - v)$. But then it follows that v is part of a component which is comprised of even components in $C - v$. Adding v it follows that v must be in an odd component $odd(C - u)$ which by assumption is unique. ☺

Note that if the unique optimal matching is not perfect we can transform the game such that the unique optimal matching will be perfect and the core has not changed. To do so add match value zero edges between all players. Note that, two players, i, j, are feasible if one is single and $d_i^t + d_j^t = w_{ij}$, thus in particular they are feasible if $w_{ij} = 0$ and the former conditions hold. Clearly the core does not change through these operations. Suppose from now that the optimal matching is perfect.

Proof of Lemma 21. First, suppose there exists a player i such that he has no tight edge, that is for all j $d_i^t + d_j^t > w_{ij}$. The expected time until i is selected is $O(N)$. If the condition still holds he reduces his aspiration level and no other player increases his aspiration level, hence D has decreased.

Now suppose that for all players i, there exists j such that $d_i^t + d_j^t = w_{ij}$. Let a *tight component*, C^t, be a maximal subgraph of players (nodes) connected by tight ($d_i^t + d_j^t = w_{ij}$) edges. Let $C_1^t, C_2^t, ..., C_k^t$ be all of the tight components. By the genericity assumption the tight components are trees, for otherwise there exists an even cycle such that two alternating partitions have the same sum of match values (compare Definition 16).

We shall show that either there exists a perfect matching (a matching such that each player is matched) within each component or there exists at least one component such that the sum of aspiration levels within the component can be reduced independently of the other components.

First, suppose that all components have a perfect matching. It follows that $D^t = \nu(V)$ and we are done.

Now, suppose that there exists a component C^t which has no perfect matching. In the following we describe a series of 'operations' that yield a reduction in the aspiration levels. We shall describe the dynamics and the expected waiting time later.

Suppose, $\#\{i \in C^t : i \in \mathcal{F}\} \neq \#\{i \in C^t : i \in \mathcal{W}\}$, that is, there are either more workers or more firms. A reduction of aspiration levels by δ by the players on the larger side and an increase by δ by the players on the smaller side yields a state where D has decreased. Note that after the increases and reductions $\mathbf{d}$ is still good, given the reductions are only δ.

Next, suppose $\#\{i \in C^t : i \in \mathcal{F}\} = \#\{i \in C^t : i \in \mathcal{W}\}$. By Lemma 26 a tree C has a perfect matching if and only if for every vertex v, the graph $C - v$ has *exactly* one odd sub-component. Hence, there exists a vertex v such that $C^t - v$ has at least 3 odd sub-components. (Given that the number of firms is equal to the number of workers there must be an odd number of odd sub-components. Thus we can dispense with the case of zero or two odd sub-components.) Without loss of generality, let v be a firm. Denote the odd sub-components resulting from $C - v$ by $C^t_{odd1}, C^t_{odd2}, \ldots$. In each of these sub-components there are either strictly more firms or strictly more workers. In particular given there are at least 3 odd sub-components at least one of them contains more firms than workers. For the latter sub-component aspiration level reductions by δ by all firms and aspiration level increases by δ by all workers yield a decrease of D, while $\mathbf{d}$ remains good.

It is important to note that these adjustments can only take place if the players on the reducing side initially have positive aspiration levels. This might not be the case for all such constructions, but it will always be the case for at least one. This follows by the proof in chapter 3.5. They show that the core can be reached by only following transitions as described above or by rematchings within tight component with non-changing aspiration levels. But note that in our model rematchings are not necessary to make transitions as described above possible. Hence, by the proof of Theorem 12 in chapter 3.5, if there is no tight component allowing an aspiration level reduction (that is the larger side of a component having all positive aspiration levels), the sum of aspiration levels is optimal, that is $D = \nu(N)$.

We shall now describe how the transitions yielding a reduction in D take place within a (sub-)component. Consider a (sub-)component such that $|\mathcal{F}| > |\mathcal{W}|$ and all firms have positive aspiration levels (the analysis is analogous for the case where there are more workers than firms). By a straightforward argument one sees that there exists at least one firm i which constitutes a leaf of the tree (i.e., only has one edge, say towards worker j). Suppose we are in a phase of positive market sentiment. Then, in expectation, after N steps worker j is activated and gets matched with firm i and receives an additional δ. (Note that j may be matched with any firm such that he receives an additional δ; we would just follow a different path.) In the meantime, given that workers are currently not price flexible (market sentiment is positive) no worker has reduced his aspiration level, hence it still holds that firm i has no other tight edge. Players who are not at a leaf in the component have two tight edges and hence no renegotiations can take place. In a next step there is a 'new leaf firm' which can meet its corresponding worker and give up δ. (Since it is a period of positive market sentiment and thus workers are currently not price flexible such transitions cannot be reversed.) In total this can happen at most N times. Hence after $O(N^2)$ iterations the sum of aspiration levels is reduced by δ when the component has more firms than workers. However, it may be that the only (sub-)component without a perfect matching has more workers than firms. In this case we may have to wait $O(N^2)$ steps until market sentiment changes and above transitions can take place. It follows that the total time until D is reduced is bounded above by $O(N^2)$. ☺

Proof of Lemma 22. A crude bound for the maximal aspiration level of players is given by w^*. In the extreme case each player has to reduce his aspiration level w^*/δ times. Thus the number of necessary aspiration level reductions is of the order of $N \cdot \frac{w^*}{\delta}$. The result now follows from Lemma 21. ☺

Proof of Lemma 23. Let M be the optimal matching. Suppose there exists $(i,j) \in M$, such that $d_i^t + d_j^t > w_{ij}$. By Lemma 19 we know that (in particular) for all $(i',j') \in M$, $d_{i'}^t + d_{j'}^t \geq w_{i'j'}$. Hence

$$
\begin{aligned}
D = \sum_{(i',j') \in M} d_{i'}^t + d_{j'}^t &= d_i^t + d_j^t + \sum_{(i',j') \neq (i,j),\ (i',j') \in M} d_{i'}^t + d_{j'}^t && (4.7)\\
&\geq d_i^t + d_j^t + \sum_{(i',j') \neq (i,j),\ (i',j') \in M} w_{i'j'} && (4.8)\\
&> w_{ij} + \sum_{(i',j') \neq (i,j),\ (i',j') \in M} w_{i'j'} && (4.9)\\
&= \sum_{(i',j') \in M} w_{i'j'} && (4.10)\\
&= \nu(V). && (4.11)
\end{aligned}
$$

This constitutes a contradiction to the assumption that $D = \nu(V)$ and the result follows. ☺

In order to prove Lemma 24 we first provide a lemma concerning the expected meeting time of two random walks on two, possibly different, sub-trees of a given tree. Our proof technique mirrors Tetali and Winkler (1991) who give a bound for the meeting time of two random walks on (the same) general graph. They show that the expected meeting time on a general graph on N vertices is at most $\frac{8}{27}N^3$.

Lemma 27. *Let T be an undirected tree on N vertices. Let two tokens x, y be placed on two vertices at time $s = 0$. Suppose that at each time step $s = 1, 2, 3, \ldots$ one of the two tokens is activated (which token is chosen may be governed by any probability distribution). The token performs a simple random walk to a neighboring vertex on a truncated tree $T_x^s \subset T$ (or $T_y^s \subset T$). Then,* *regardless of the rule of activation, if in every time step the (unique) path between x and y is in T_x^s (respectively T_y^s) the expected time $M_{truncated}(x, y)$ until the tokens meet is $O(N^2)$.*

In order to prove the lemma we shall need the following two lemmas from Tetali and Winkler (1991) which can be proved by a straight forward calculation due to the reversibility of the Markov chain for random walks on undirected graphs. Given a connected graph T, let $H_T(x, y)$ be the hitting time from x to y on T, that is the expected time until y is first reached when starting a simple random walk at x on the graph T.

Lemma 28. *(Tetali and Winkler 1991, Lemma 2) Let x, y and z be vertices of a connected undirected graph T. Then*

$$H_T(x, y) + H_T(y, z) + H_T(z, x) = H_T(x, z) + H_T(z, y) + H_T(y, x) \qquad (4.12)$$

Lemma 29. *(Tetali and Winkler 1991, Lemma 3) On any graph T the vertex relation given by*

$$x \leq y \iff H_T(x, y) \leq H_T(y, x) \qquad (4.13)$$

is transitive.

It immediately follows that there exists a vertex v which is *minimal* in this order. Call such a vertex *remote* and note that we have $H_T(x, v) \leq H_T(v, x)$ for every vertex $x \in T$.

Proof of Lemma 27. The proof mirrors Tetali and Winkler (1991). We shall show that $\forall x, y \in T$

$$M_{truncated}(x, y) \leq H_T(x, y) + H_T(y, v) - H_T(v, y). \qquad (4.14)$$

where v is a remote vertex. Define a potential function

$$P(x,y) = H_T(x,y) + \overbrace{H_T(y,v) - H_T(v,y)}^{\geq 0 \text{ since } v \text{ is remote}} \quad (4.15)$$

$$\overset{\text{Lemma 28}}{=} H_T(y,x) + H_T(x,v) - H_T(v,x). \quad (4.16)$$

Hence, independent of which token is moved we have

$$P(x,y) = 1 + P(\bar{x},y) = 1 + P(x,\bar{y}), \quad (4.17)$$

where for a function f, $f(\bar{x})$ is defined to be the average of $f(y)$ over all neighbors y of x (note that this defines an expectation). This is the case since, independent of the token moved, the expected hitting time decreases by one for the given token.

Define $\beta(x,y) = M_{truncated}(x,y) - P(x,y)$. Among all pairs maximizing $\beta(\cdot,\cdot)$ choose x^*, y^* at minimal distance from each other. Now suppose inequality (4.14) does not hold for all $x, y \in T$, that is $\beta(x^*,y^*) > 0$ (note that $x^* \neq y^*$ follows, for otherwise $M_{truncated}(x^*,y^*) = 0 \leq P(x^*,y^*)$). W.l.o.g. assume that the token moved next is on x^*. Then

$$M_{truncated}(x^*,y^*) = P(x^*,y^*) + \beta(x^*,y^*) \quad (4.18)$$

$$= 1 + P(\bar{x^*},y^*) + \beta(x^*,y^*) \quad (4.19)$$

$$= 1 + \frac{1}{d(x^*)} \sum_{x_i : x_i \text{ neighbor of } x^*} P(x_i,y^*) + \beta(x^*,y^*) \quad (4.20)$$

$$> 1 + \frac{1}{d(x^*)} \sum_{x_i : x_i \text{ neighbor of } x^*} P(x_i,y^*) + \beta(x_i,y^*) \quad (4.21)$$

$$= 1 + \frac{1}{d(x^*)} \sum_{x_i : x_i \text{ neighbor of } x^*} M_{truncated}(x_i,y^*) \quad (4.22)$$

$$= M_{truncated}(x^*,y^*) \quad (4.23)$$

where $d(x)$ is the degree of node x in T. Inequality (4.21) holds since for all i,

$\beta(x_i, y^*) \leq \beta(x^*, y^*)$ and there exists at least one j such that x_j is closer to y^* than x^* and thus $\beta(x_j, y^*) < \beta(x^*, y^*)$ (since we assumed x^*, y^* to be at minimal distance maximizing $\beta(\cdot, \cdot)$). Equation (4.22) holds by a similar argument as for equation (4.17). But $M_{truncated}(x^*, y^*) > M_{truncated}(x^*, y^*)$ constitutes a contradiction and thus proves inequality (4.14). By Georgakopoulos and Wagner (2013, Corollary 8) the hitting time on a tree is bounded above by the square of the number of edges $(N-1)$. Thus we have

$$M_{truncated}(x, y) \leq H_T(x, y) + H_T(y, v) = O(N^2). \tag{4.24}$$

☺

Lemma 30. *Let T be a tight component (a tree) that allows a perfect matching. Suppose there are exactly two remaining singles i and j. Then after a finite number of transitions where i remained single throughout i matches with some j'.*

Proof. We shall show that the path in the component between i and j must be alternating between unmatched and matched edges. Suppose, by contradiction, that the path between i and j is not alternating between unmatched and matched edges. If i and j share an edge we are done. Else, let i^* be the first firm on the (unique) path from i to j such that i^* is not matched to a worker on the path. (Note that i^* is matched elsewhere.) Let j^* be the worker preceding i^* on the path. We shall construct an instance of Lemma 26. Delete i^* from T. Then the component including i is such that all matches are within the component (since i^* is matched elsewhere).

Since this component does not contain j and contains i and otherwise only contains matched pairs it is of odd size. On the other hand let j^{**} be matched with i^*. Now consider the component containing j^{**} in $T - i^*$. Since j^{**} is matched to i^* it also must be odd. This contradicts Lemma 26 and, by contradiction, establishes the claim that there is an alternating path of unmatched and matched edges between i and j.

Let the unique path be $i, j_1, i_1, j_2, i_2, \ldots, j_k, i_k, j$. We can now construct a path, proving the lemma. Suppose i is activated and matches with j_1, leaving i_1 single. Next i_1 is activated and matches with j_2, leaving i_2 single. Continuing in this manner we see that after k steps i_k is activated and matches with j. ☺

Proof of Lemma 24. Given $d_i^t + d_j^t \geq w_{ij}$ for all i, j, rematches only occur when at least one player is single. Thus the number of matches is weakly increasing over time and it suffices to bound the expected time until two singles match. It follows immediately from Lemma 23 that each tight component allows a perfect matching. We shall compute the expected time until the number of singles is zero.

We describe the transitions without taking market sentiment into account since it would only accelerate the process. Note that, since there may be players with aspiration level zero, we cannot necessarily follow a transition pattern as described in the proof of Lemma 21.

We shall construct two simple random walks which we then relate to our process. Let X, Y be two simple random walks on truncated subgraphs of T as defined in Lemma 27 and let $s = 0, 1, 2, 3, ...$ be the time steps at which one of the random walks moves on some truncated tree (such that the path between the two random walks is on the truncated tree). Let the rule of activation for the random walks be such that both random walks always move twice consecutively. That is, for all s odd, if X (respectively Y) is activated in s, then X (respectively Y) is also activated in $s+1$. Define for X (analogously for Y) the following time steps. Initialize $T_0^X = T_0^Y = 0$, then for $i = 1, 2, 3, \ldots$

let

$$T_i^X = \inf\{s > T_{i-1} : s \text{ odd, and } X_s \neq X_{s-2}\}, \tag{4.25}$$

and

$$\widetilde{T} = \inf\{s : X_s = Y_s\}. \tag{4.26}$$

The sequence of time steps at which the random walk X (or Y) has moved away from odd s after two time steps or has collided with Y (respectively X) is given by

$$\mathcal{T}^X = (T_1^X, T_2^X, \ldots, T_I^X, \widetilde{T}), \tag{4.27}$$

$$\mathcal{T}^Y = (T_1^Y, T_2^Y, \ldots, T_J^Y, \widetilde{T}), \tag{4.28}$$

where I, J are two random variables describing the lengths of the sequences. Note that for all time-steps $s \notin \mathcal{T}^X \cup \mathcal{T}^Y$ the random walk returned to its previous place after two steps. Define $\widetilde{X}$ and $\widetilde{Y}$ to be the (not simple) random walks of step-size two on truncated trees. Until the two underlying random walks X and Y collide, $\widetilde{X}$ and $\widetilde{Y}$ move at times $\mathcal{T}^X \cup \mathcal{T}^Y$.

We now relate the two random walks, $\widetilde{X}, \widetilde{Y}$, to our process. We shall first consider the case where only two singles remain in a given tight component. Suppose i and j are the only two remaining singles. Note that the two singles will always remain in the *same* tight component. Now consider those time steps where a current single becomes matched and a previously matched player becomes single. (The expected interval between such time steps is upper bounded by $O(N)$.)

Suppose we start the random walk $\widetilde{X}$ on the single firm i and the random walk $\widetilde{Y}$ on the single worker j. Suppose i (or j) gets activated in a given time step. Then $\widetilde{X}$ (or $\widetilde{Y}$) is also activated and moves to a node in the tree at distance two from i, say i' (via j') such that (i', j') were matched in the previous period. This is the condition for the truncated tree on which the random walk $\widetilde{X}$ (or $\widetilde{Y}$) moves, namely the truncated tree starting at i only has branches alternating between tight unmatched and matched edges. Any such node i' is picked uniformly at random amongst the matches of current neighbors. Thus $\widetilde{X}$ (respectively $\widetilde{Y}$) perform the previously defined two-step random walk. The time steps at which this happens are precisely given by $\mathcal{T}^X$ (respectively $\mathcal{T}^Y$). This indicates that player i is matching with player j'. Now if $j' = j$ the two last remaining singles match and the simple random walks X and Y collide. Else i' is now occupied by $\widetilde{X}$.

We are left to give an upper bound for $\mathbb{E}[I + J + 1]$. First note that we have $\mathbb{E}[I + J + 1] < M_{truncated}(X, Y)$. We shall now argue that

$$\mathbb{E}[I + J + 1] = \frac{1}{4} M_{truncated}(X, Y) \tag{4.29}$$

and thus the bound we shall give is sharp. If X (or Y) is moved in periods $s+1$ (odd) and $s+2$ we have $\mathbb{P}(X_{s+2} = X_s) = 1/2$. This is the case since on the truncated tree the random walk first moves along a tight edge and then necessarily moves back or along a matched edge. Thus $\mathbb{E}[T^X_{i+1} - T^X_i] = 4$ given the expectation of the geometric distribution with parameter $1/2$. Now the result follows.

Overall, given the singles need to be activated in order for the random walks to be moved, the meeting of the two single tokens takes $O(N^3)$ iterations.

Now suppose there are $2k$ ($k \in \{1, \dots, N\}$) remaining singles. The expected time until there are only $2(k-1)$ singles is bounded by $O(N^3)$. Then for $k = O(N)$ we have a total expected time until the core is reached of $O(N^4)$.[10]

[10] In line with Tetali and Winkler (1991, Conjecture 2) we conjecture that this bound is not optimal and the total meeting time of k random walks on truncated trees is of the same order as the two-token meeting time. Together with the necessary activation of single tokens this would yield the bound $O(N^3)$.

Finally we have to show that any core state is absorbing. We have for all $i \in V$ that i is matched and for all i, j that $d_i^t + d_j^t \geq w_{ij}$ since we are in a core state with the unique perfect matching. Thus there are no feasible pairs of players and the dynamic has reached a fixpoint. ☺

Proof of Lemma 25. Given that i is activated, suppose there exists a feasible match. The calculation whether a player has no better alternative requires $O(N)$ for each $j' = 1, ..., N$, and thus $O(N^2)$ in total. The case where no feasible match exists for i requires at most $O(N)$ operations. The result now follows. ☺

Proof of Theorem 17. Combining Lemma 20, 22, and 24, the expected maximum number of steps to absorption is $O(N^3 \cdot \max\{\log\log N, \frac{w^*}{\delta}\})$ proving the first part of the theorem. By Lemma 25 we know the complexity of each iteration, $O(N^2)$, and the second part of the theorem follows. ☺

4.6 Conclusion

In this chapter we have shown that agents in large decentralized assignment markets can learn to play core outcomes in polynomial time through simple aspiration adaptation learning rules. Players have no knowledge about other players' past payoffs or actions, their distribution of preferences, or about the value of different matches. Players are randomly activated and thus update their action in an asynchronous, random manner. Players find a feasible match if one exists and the price they agree to is influenced by market sentiment. While it was previously shown that a variant of this random process is absorbed in the core in finite time with probability one (Chen et al. 2011, Nax and

Pradelski 2014, Biró et al. 2012, Klaus and Payot 2013) the dynamic described here converges in polynomial time.

Our result is in the same spirit as Bayati et al. (2005, 2014), Salez and Shah (2009), Celis et al. (2010), Azar et al. (2009), Ackermann et al. (2011), and Kanoria et al. (2011), all of whom study the rate of convergence of dynamics for the assignment game. There are two fundamental differences however. First, information decentralization as defined in the introduction is not satisfied; players know the reservation values of other players. Second, all positive results (that is polynomial convergence time) are only shown for deterministic processes with a fixed order of activation, violating our definition of market decentralization. We are the first to show polynomial time convergence for a stochastic decentralized dynamic for the assignment game.

Market sentiment turns out to be crucial in our analysis. We have shown that without market sentiment our dynamic may take exponentially long to converge. This, alongside according results for NTU models (Ackermann et al. 2011) support the conjecture that it is impossible to devise a pure better response dynamic which converges in polynomial time. It is an open question whether conditions for polynomial time convergence for general classes of games can be formulated. Our proof technique regarding the duality with random walks and the use of market sentiment may be pointers towards a possible path of investigation for other games on networks.

Chapter 5

The dynamics of social influence

5.1 Introduction

Social influence describes a process in which individual opinions or behaviors are affected by the opinions or behaviors of others in a group. In this chapter we study a general family of dynamic models of social influence. Agents update their action at random points in time. Their decisions are influenced by the actions previously taken by other members of the society, where the strength of this influence varies across agents. We distinguish between two cases: On the one hand, we consider the case where social influence arises from responding to the *adoption level* (*Adoption*). On the other hand, we consider the case where social influence arises from the *recent usage* (*Usage*). In *Adoption* each player's action is only counted once. In *Usage* each player's action is counted by the number of periods in which he took that action. Thus the repeated action by a single player may be counted several times. We first identify the equilibria of these models and then study their stability in stochastic environments. We characterize the stable states of the dynamics and find that the behavior differs for the two models.

The discussion of herding behavior and social influence has a long history. For example Trotter (1916) identifies herd instinct as one of the primary instincts along with self-preservation, nutrition, and reproduction. Freud (1921a,b) points to an individual's tendency to follow the masses ("Herabsinken zum Massenindividuum"). For other early discussions of social influence see, for example, Le Bon (1895). Subsequently the study of the influence of a group on the decision or action of an agent entered the disciplines of economics, sociology, and psychology (see, for example, Keynes 1930, 1936, Hamilton 1971, Schelling 1971, Shiller 1984, 2000).

In the next section we discuss the related literature. In section 5.3 we introduce the model and the different types of social influence. Section 5.4 contains all the results. After some initial observations we introduce our guiding example. We then analyze the social influence models *Adoption* and *Usage* and compare them. Section 5.5 concludes.

5.2 Related literature

Before discussing the literature on social influence in more detail we shall make an important remark regarding the difference between social learning and social influence (see, for example, Montgomery and Casterline 1996, Young 2009a). *Social learning* takes into account other players' signals and information in evaluating alternative actions.[1] *Social influence* "refers to the effects of interpersonal interactions that derive their power from factors that are intrinsically 'social' "(Montgomery and Casterline 1996).[2] In this chapter, we study the dynamics of social influence. Nevertheless, *Usage* is also a key ingredient of social

[1] See, for example, Banerjee (1992) and Ellison and Fudenberg (1995).

[2] This could be the desire to conform, avoid conflict, mitigate uncertainty, or a more direct utility from a shared use of an innovation such as mobile phones.

learning models. The cumulative usage of, for example, an innovation may be a 'real' signal of a utility advantage (see Young 2009a). For example, if a taxi driver frequently uses a mobile phone application this is likely an indicator (to other drivers) that it actually helps him to gain more work.[3]

Models of social influence have numerous applications, for example, political and social movements (Schelling 1978, Lohmann 1994, Cabinet Office 2012), diffusion processes such as innovation adoption (Rogers 1962, Bass 1969, Valente 1996, Meade and Islam 2006, Young 2009a), and financial herding (e.g., Scharfstein and Stein 1990, Welch 1992, Shiller 2000). There is broad experimental evidence for social influence. Asch (1955) conducted a series of enlightening experiments showing that a considerable proportion of subjects trust the majority over their own senses. More recently Salganik et al. (2006) show the effects of social influence in a study on music taste. Other experimental studies include voting and opinion polls (Cukierman 1991), human fertility (Bongaarts and Watkins 1996), diffusion of information technologies (Teng et al. 2002), stock market participation (Hong et al. 2004), household energy consumption (Schultz et al. 2007), mobile phones (de Silva et al. 2011), and mobile banking (Yu 2012).

Considering the above examples there are cases where social influences takes the form of *Adoption* and others where it takes the form of *Usage*. We shall use innovation adoption as our guiding motivation and example.[4] For example, consider the spread of mobile phones and its associated applications in developing countries.[5] On the one hand, from a 'hardware perspective' the act of buying a mobile phone is a one-off act and thus the relevant information

[3]For example, HAILO was the first provider of a mobile phone application for black cabs in London.

[4]'Innovation' and 'new product' are used interchangeably in the literature.

[5]See, for example, Kiringai and Fengler (2010) for a discussion of mobile money in Kenya. They find for both, mobile phones and mobile money an S-curved adoption curve supporting the claim that social influence is present (see Young 2009a).

to other potential buyers is the number of current adopters (*Adoption*). That is, when considering which phone to buy you might look at the phones your friends currently use but it is unlikely that you take into account the phones they previously used or how often they have switched. On the other hand, from a 'software perspective' (e.g., mobile money, WhatsApp) the cumulative usage is likely to be a more important driver of social influence than the mere installation of the product (*Usage*).[6] That is, when considering competing services to send text messages you likely take into account the cumulative usage of their friends rather than which services are merely installed on their phones.

The key novelty of our framework is that we differentiate between social influence arising from responding to the number of adopters versus social influence arising from responding to the cumulative usage. We consider noisy best-reply models and show a selection within the set of fixpoints (independent of the starting state). Interestingly the two seemingly similar models often exhibit different long-run behavior.

Our model is most closely related to Schelling (1978, Chapter 3) and Granovetter (1978). Schelling (1978, Chapter 3) describes the class of *critical mass models* of social interaction and gives several instructive examples. He notes that "though perhaps not in physical and chemical reactions, in social reactions it is typically the case that the 'critical number' for one person differs from another's" (Schelling 1978). He goes on to define the *tipping value* which, for each player, determines the critical mass of the aggregate action of the population for which a player will 'tip over' from playing one action to another. The heterogeneity of social influence has been confirmed in recent work,

[6]Note that this differentiation has not gone unnoticed in marketing departments. Hardware providers such as Apple focus their reporting and marketing effort on the number of products sold, i.e., roughly the number of adopters. On the other hand, software providers, such as Skype, focus their analysis and messaging on the number of people who are online (or, more generally, use the product) at any point in time, i.e., the usage intensity.

suggesting that cognitive factors influence the propensity to herding behavior (Dohmen et al. 2012, Baddeley et al. 2007, Moussaid et al. 2013). In a similar vein Granovetter (1978) develops a threshold model for binary decision games and describes the nature of fixpoints given the individual players' thresholds or tipping values.

We shall adopt the assumption of heterogeneous players who repeatedly revise their action which is binary. As in Schelling (1978) and Granovetter (1978) we employ a threshold model with heterogeneous tipping values. We consider a continuous time model of asynchronous updating, thus guaranteeing the existence of equilibria (see Lopez-Pintado and Watts 2008). Granovetter (1978) notes that if misperception is random "the situation is more complex and the stability of the underlying equilibrium becomes particularly important". The study of this scenario is precisely the purpose of the current chapter.

For further studies on social influence, see Macy (1991), Dodds and Watts (2004, 2005), Horst and Scheinkman (2006), Bramoulle (2007), Lopez-Pintado and Watts (2008), Scheinkman (2008), Baddeley (2010), Young (2011), Moussaid et al. (2013), Babichenko (2013). Another strand of the literature termed *herding* analyzes social influence under the assumption of Bayesian learning (Banerjee 1992, Bikhchandani et al. 1992, Scharfstein and Stein 1990, Welch 1992, Devenow and Welch 1996, Cont and Bouchaud 2000, Bouchaud 2013). These models are often applied in finance and explain with the help of information cascades the build up of bubbles. The literature on the evolution of social norms is also closely related (see, for example, Axelrod 1986, Young 1993a, 1998). Further, the spread of infectious diseases in biology is often modeled in a similar way, but the models considered are mainly independent interaction models (see Kermack and McKendrick 1927 and Dodds and Watts 2005).[7]

[7]In an independent interaction model the probability of infection is independent of the

As in much of the literature we consider a 'mean-field' approach, assuming that an agent observes every other agent with equal probability. A separate strand of the literature studies social influence given an underlying network of interactions (see, for example, Blume 1993, Valente 1996, Young 1998, Morris 2000, Jackson and Watts 2002, Centola and Macy 2007, Golub and Jackson 2010, Banerjee et al. 2013). The processes we consider are Markovian. We use the concept of stochastic stability (see section 1.3). The idea is to study a perturbed version of the original process, such that the resulting Markov process is irreducible and ergodic and therefore the process has a unique stationary distribution. By letting the level of noise approach zero one can identify those states that will be observed in the long-run with arbitrarily high probability. We also make use of recent work on reinforced random walks (Pinsky 2013). Pinsky analyzes a random walk on $\mathbb{Z}$ whose probability of moving left or right depends on the recent history. By an appropriate translation we can use his results to find the long-run stable states of our dynamic.

5.3 The model

We shall first introduce the general framework for analyzing social influence. Let $N = \{1, ..., n\}$, $n \in \mathbb{N}$ be a finite set of types. Let $P = \{1, ..., p\}$, $p \in \mathbb{N}$ be the set of players where each player is associated with exactly one type.[8] Let $q_i = \frac{\sum_{j=1}^{p} \mathbf{1}_{j \text{ is of type } i}}{p}$ (for $i = 1, \dots, n$) be the ratio of players of type i. Let

number of contacts with infected agents. On the other hand in threshold models (as considered here) the probability of 'infection' depends on the number of currently 'infected' agents.

[8]More generally, we could think of a person comprising of several players, possibly of different types. This could be the case for multiple reasons. For example, one person might be more important (or observable) in the population than another person and therefore his adoption or usage might be 'counted' multiple times. Further, a person might simply use a product more than another person, or use a product more or less dependent on which type currently would like to use the product.

$A = \{m, d\}$[9] be the actions available to each player $i \in P$.

Let $u_i : [0,1] \times A \to \mathbb{R}$ be the utility of agent $i \in P$ when observing the signal about the society $s \in [0,1]$ and playing action $a \in A$. Suppose that the utility of an action is separable into a component arising from a player's inherent preference for an action and a component specifying the utility he derives from social conformity. After normalizing, let $p_i \in \mathbb{R}$ be player i's direct utility difference when playing action d over action m. Further let $\rho_i \in \mathbb{R}^+$ be a player's *index of social conformity*. Finally, suppose that the impact of social influence is linear. A player's utility from playing action a is now given by

$$u_i(a) = \begin{cases} p_i + \rho_i s & \text{if } a = d, \\ \rho_i(1 - s) & \text{if } a = m. \end{cases} \tag{5.1}$$

Note that this is essentially a coordination game where player have heterogeneous mixed strategies.

Let $f_i : \mathbb{R}^2 \to [0,1]$ be the *response function* for player i, specifying the probability to play action d given his utilities $u_i(s,d) \in \mathbb{R}$ and $u_i(s,m) \in \mathbb{R}$. Note that $1 - f_i(\cdot,\cdot)$ is the probability that i plays action m. We shall initially consider a best-response model:

$$f_i = \begin{cases} 1 & \text{if } u_i(s,d) > u_i(s,m), \\ 0.5 & \text{if } u_i(s,d) = u_i(s,m), \\ 0 & \text{else.} \end{cases} \tag{5.2}$$

We shall consider a continuous time process where each player is activated by iid Poisson arrival processes. Let $t = 0, 1, 2, \ldots$ be the time steps in which a

[9] *Dextre* and *maldextre* are the terms for right and left in Esperanto.

(unique) player gets activated.[10,11] In a given time step t the activated player i will be called *active*. Define

$$\mathbf{1}_i^t = \begin{cases} 1 & i \text{ is active in } t, \\ 0 & \text{else.} \end{cases} \tag{5.3}$$

Let $s(t) \in [0,1]$ be agent i's observation of society at time t. For each player i, let a_i^t be the action he takes at time t. The state at the end of a given period t is given by the action profile $\mathbf{a}^t = (a_i^t)_{i \in P}$. Let $\mathbf{a}^0$ be any permissible initial configuration. Then

$$a_i^t = \mathbf{1}_i^t \cdot B^t[f_i(u_i(s(t), d), u_i(s(t), m))] + (1 - \mathbf{1}_i^t) \cdot a_i^{t-1} \tag{5.4}$$

for all $t \geq 1$, where $(B^t)_{t \in \mathbb{N}}$ is a family of independent Bernoulli random variables taking values in A.

Let $\bar{a}^t = \sum_{i=1}^p \mathbf{1}_{a_i^t = d}/p \in [0, \frac{1}{p}, \ldots, 1]$ be the *population's average action* in period t. We consider two models of social influence:

- **Adoption level.** An active player responds to the adoption level (in short *Adoption*):

$$s^{adoption}(t) = \bar{a}^{t-1} = \frac{\sum_{j \in P} \mathbf{1}_{a_j^{t-1} = d}}{p} \tag{5.5}$$

- **Recent usage.** An active player responds to the recent usage in the past k periods (in short *Usage*):

$$s^{usage}(t) = \frac{\sum_{v=t_0}^{t-1} \sum_{i=1}^p \mathbf{1}_{a_i^v = d} \mathbf{1}_i^v}{k} \quad \text{for} \quad t_0 = t - k \tag{5.6}$$

[10] Note that by the assumption of iid Poisson arrival processes we have that almost surely no two players are activated at the same moment.

[11] When unambiguous we shall sometimes omit the specification of the time period.

When unambiguous we shall write $s(t)$ for $s^{adoption}(t)$ or $s^{usage}(t)$ respectively.[12,13] Note that both $s^{adoption}(t)$ and $s^{usage}(t)$ include any given player's action. This is reasonable when players are presented with the aggregate statistic, but the analysis also carries through if one excludes a given player's action.

To illustrate, suppose there are four players. Player 1 initially plays d and all other players play m. Suppose we are in time step $t = 7$ and play unfolded as shown in table 5.1. For *Adoption* the relevant observation in period $t = 7$

Table 5.1: Actions up to $t = 6$

time step	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
active player	–	player 2	player 1	player 4	player 1	player 2	player 2
player 1	d		$\boxed{d}$		$\boxed{d}$		
player 2	m	d				$\boxed{d}$	$\boxed{d}$
player 3	m						
player 4	m			$\boxed{m}$			

Relevant actions for $s^{adoption}(7)$ circled, relevant actions for $s^{usage}(7)$ boxed ($k = 5$).

of current adopters of action d is $s^{adoption}(7) = 50\%$ (the relevant actions are circled in Table 5.1). For *Usage* suppose, for example, that $k = 5$. Then the observation in period $t = 7$ of the cumulative usage of action d is $s^{usage}(7) = 80\%$ (the relevant actions are boxed in Table 5.1). Note that in our example player 3's previous actions have no influence on the observed cumulative usage.

We shall analyze the two models for several regimes of sampling and errors. Initially we will study the unperturbed dynamics as defined above. That is, in the case of *Usage*, $k = t$. We then turn to study perturbed dynamics. For *Usage* the natural perturbation is to limit the 'look-back' window to a constant k. For the *Adoption* regime we shall consider a uniform action tremble. That

[12] For *Usage* we generally assume that k is fixed. The only exception we shall consider is $k = t$, that is, the sample comprises all actions taken in the past.

[13] For *Usage*, to be precise we need to define $s(t)$ differently for $t < k$. We can simply assume the average of the past t actions.

is, there exists a small probability $\varepsilon > 0$ such that a player picks an action uniformly at random. The response function for player i then is

$$f_i = \begin{cases} 1 - \frac{\varepsilon}{2} & \text{if } u_i(s(t), d) > u_i(s(t), m), \\ 0.5 & \text{if } u_i(s(t), d) = u_i(s(t), m), \\ \frac{\varepsilon}{2} & \text{else.} \end{cases} \tag{5.7}$$

5.4 Analysis

Definition 31. *Let each player i's* tipping value $\mu_i \in \mathbb{R}$ *be such that he wants to play d if*

$$\mu_i < s(t), \tag{5.8}$$

and he wants to play m if

$$\mu_i > s(t), \tag{5.9}$$

and if $\mu_i = s(t)$ he is indifferent. Note that μ_i is the (unique) zero of the function $u_i(s(t), d) - u_i(s(t), m)$.

It immediately follows that a player with $\mu_i \in (-\infty, 0)$ always prefers to play action d and a player with $\mu_i \in (1, \infty)$ always prefers action m. Note that players of the same type have the same tipping value.

We shall make the simplifying assumption that for all players i, μ_i is not a multiple of $1/p$. Thus a player always has a unique best response. Note that this assumption is generic if the parameters of the utility functions are drawn from continuous independent distributions.

Definition 32. *Let Agg be the* Aggregate Dynamic*. That is, given the population's average action $\bar{a}$, Agg gives the share of players who would play d,* *given they observe this state.*

$$Agg : [0,1] \to \left\{0, \frac{1}{p}, \frac{2}{p}, \dots, 1\right\} \tag{5.10}$$

$$Agg(\bar{a}) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}_{\mu_i < \bar{a}} \tag{5.11}$$

Lemma 33. *Agg has at least one fixpoint. If x^* is a fixpoint of Agg all players of the same type play the same action, that is*

$$x^* \in \{\bar{q}_i = \sum_{j=1}^{i} q_j : \mu_i < \bar{q}_i < \mu_{i+1}\}_{i=1,\dots,n}.^{14} \tag{5.12}$$

Proof. We shall first show that there exists a fixpoint. If $Agg(1) = 1$ then 1 is a fixpoint and similarly if $Agg(0) = 0$, 0 is a fixpoint. Thus we still have to consider the case where $0 < Agg(1) < 1$. Define the function:

$$g : [0,1] \quad \to \quad \left\{-1, ..., \frac{-1}{p}, 0, \frac{1}{p}, ..., 1\right\} \tag{5.13}$$

$$g(x) \quad \mapsto \quad Agg(x) - x \tag{5.14}$$

It suffices to prove that there exists x such that $g(x) = 0$. Since $Agg(0) > 0$ we have $g(0) \neq 0$ and thus $g(0) > 0$. Also $Agg(1) < 1$ and thus $g(1) < 0$. Hence there must eventually be a change from positive to negative sign with increasing x. Remember that g is defined on $\left\{-1, ..., \frac{-1}{p}, 0, \frac{1}{p}, ..., 1\right\}$. We have by monotonicity of Agg that for every $x < 1$, $g\left(x + \frac{1}{p}\right) \geq g(x) - \frac{1}{p}$, that is the function g decreases at most in steps of $\frac{1}{p}$. Therefore there exists x^* such that $g(x^*) = 0$ and thus $f(x^*) = x^*$.

For the second part of the lemma note that if x^* is a fixpoint of Agg all players of the same type necessarily play the same action. It remains to show that for any fixpoint $\bar{q}_i$ we have $\mu_i < \bar{q}_i < \mu_{i+1}$. By contradiction, suppose there exists a

[14] Set $\mu_{n+1} = 1$ for completeness.

fixpoint $\bar{q}_i < \mu_i$. Then for the share of players q_i it is optimal to play m when observing $\bar{q}_i$, given that their tipping value is μ_i and thus $\bar{q}_i$ is not a fixpoint. A similar argument applies for $\bar{q}_i > \mu_{i+1}$. ☺

We shall now introduce an example of innovation diffusion which will guide us throughout the chapter. Given a population as proposed by Rogers (1962), suppose there are

- 2.5% *innovators*, who always play the innovation action d, that is, they play the innovation independent of social influence and hence their tipping value is 'negative' ($\mu_{\text{innovators}} < 0$),
- 13.5% *early adopters*, who play the innovation if at least 'few' play the innovation ($\mu_{\text{early adopters}} = 9.25\%$),
- 34% *early majority*, who play the innovation if at least an 'intermediate proportion' play the innovation ($\mu_{\text{early majority}} = 32\%$),
- 34% *late majority*, who play the innovation if at least 'many' play the innovation ($\mu_{\text{late majority}} = 68\%$),
- 13.5% *laggards*, who play the innovation if at least 'almost everybody' play the innovation ($\mu_{\text{laggards}} = 90.75\%$),
- 2.5% *non-adopters*, who never play the innovation ($\mu_{\text{non-adopters}} > 1$).[15]

Note that the broad categories (i.e., few, intermediate proportion, etc.) represent a 'coarse' mapping from observations to beliefs.[16] Figure 5.1 shows the distribution of tipping values. Figure 5.2 shows the function Agg and the fixpoints $x_1^*, \ldots, x_5^*$.

[15]Note that in Rogers (1962) the two last categories are considered as one. He mentions the possibility to split them in order to create a symmetric distribution. Rogers argues that adopter distributions are well approximated by the normal distribution; for example innovators are 2 standard deviations or more above the mean level of innovativeness.

[16]In political philosophy and psychology it is understood that some beliefs may be coarse-grained while others are fine-grained (see, for example, Sturgeon 2009).

Figure 5.1: Tipping value distribution – x-axis shows μ_i, y-axis shows q_i

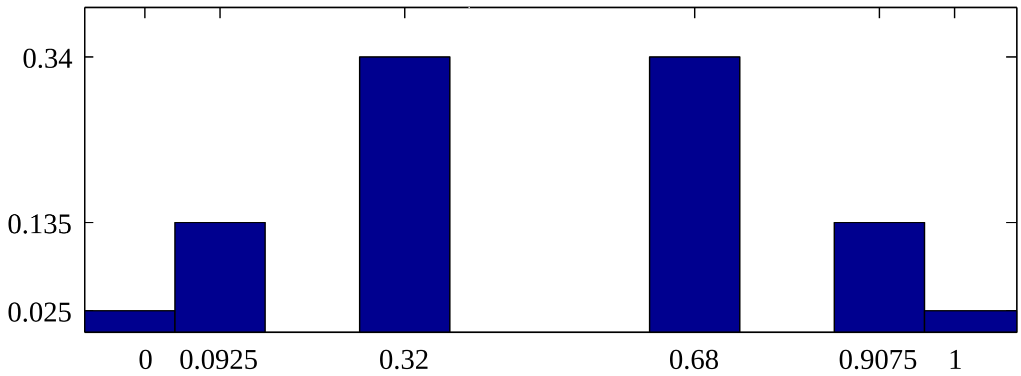
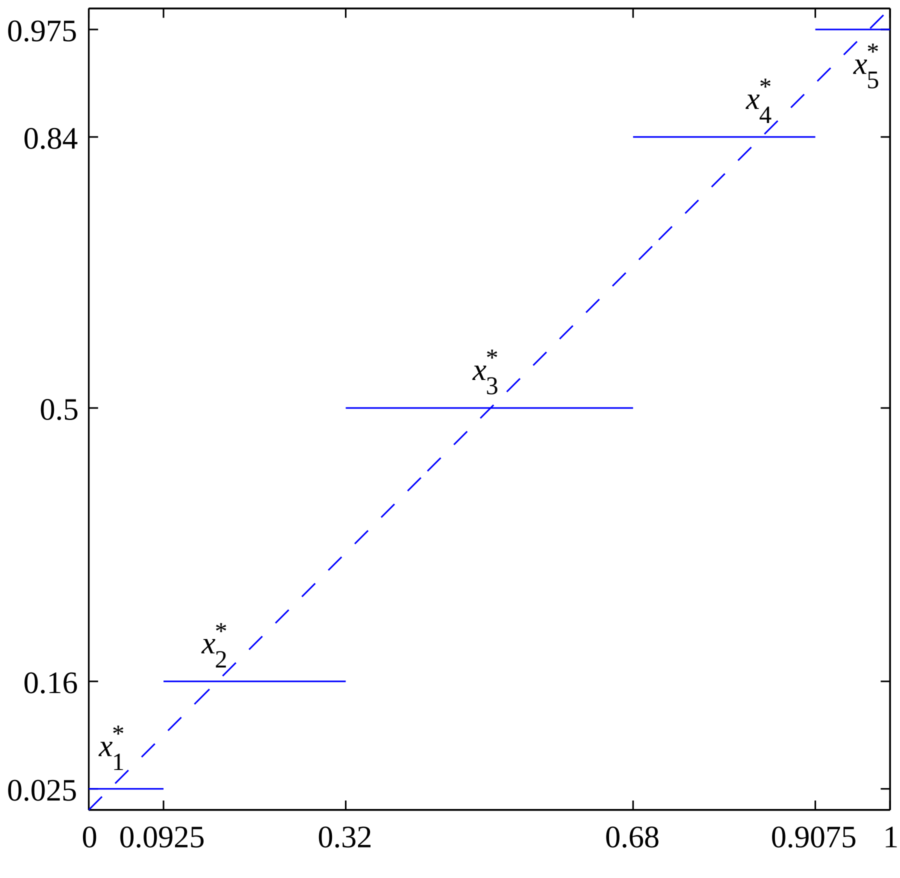


Figure 5.2: Aggregate dynamic – x-axis shows μ_i, y-axis shows Agg



5.4.1 Adoption

In this section we shall consider the social influence model *Adoption*. An active player bases his decision on the number of current adopters (see equation 5.5):

$$s(t) = \bar{a}^{t-1} = \frac{\sum_{j \in P} \mathbf{1}_{a_j^{t-1}=d}}{p}$$

Theorem 34. *The dynamic has at least one absorbing state. The absorbing states of the dynamic process coincide with the fixpoints of Agg and each absorbing state is associated with exactly one fixpoint of Agg and vice versa. The set of absorbing states which can be reached is dependent on the initial state.*

Proof. By contradiction, suppose there exists an absorbing state $\mathbf{a}^*$ with $x^* \in [0,1]$ such that x^* is not a fixpoint of Agg. Suppose $x^* > Agg(x^*)$ (the other case is analogous. That is, the number of players playing d is x^*, however the number of players for whom it is desirable to play d is $Agg(x^*)$ which is strictly smaller than x^*. Hence there exists at least one player i playing d for whom $\mu_i > x^*$ and if i is activated in a given period he will change his action from d to m. Thus the state $\mathbf{a}^*$ is not an absorbing state. Since by Lemma 33 Agg has at least one fixpoint, for a given fixpoint x^* we can assign an action to each player such that the action profile is an absorbing state. We simply order the players by their tipping values (from small to large) and assign action d successively to players until the ratio x^* is reached. Finally note that the dynamic is not ergodic and therefore the absorbing states attainable depend on the initial state $\mathbf{a}^0$. ☺

Note that Theorem 34 is closely related to the convergence result in Babichenko (2013). His model, in our language, allows for negative social influence. They show convergence to approximate Nash equilibrium if the observation of society is discretized, that is $\lfloor s^{adoption} \rfloor_\delta$ for some $\delta > 0$. Note that Babichenko (2013) also show that their dynamic converges in $O(n \log n)$ time steps.

Perturbed dynamics

We now consider the perturbed process with a uniform error rate. Consider the stochastic process governing the change of $\bar{a}$. The process is Markovian and its recurrent classes are characterized by the fixpoints of OBR. It is a regular perturbed Markov process and we can therefore use stochastic stability analysis (see section 1.3). We consider the long-run behavior of the process when ε becomes small.

The process governing the change of $\bar{a}$ has a linear transition structure, namely, to go from state $\frac{i}{p}$ to $\frac{j}{p}$ one has to pass through all the states $\frac{i+1}{p}, \ldots, \frac{j-1}{p}$ for $i < j$ and similarly through all the states $\frac{i-1}{p}, \ldots, \frac{j+1}{p}$ for $i > j$. Hence to find the least resistant paths between any two recurrent classes it suffices to calculate the resistance between any two *neighboring* recurrent classes.

Lemma 35. *The resistances of the paths between two neighboring fixpoints of Agg (recurrent classes) x_i^*, x_{i+1}^* $(x_i^* < x_{i+1}^*)$ are given by:*

$$r_{x_i^*, x_{i+1}^*} = \max_{x \in [x_i^*, x_{i+1}^*]} \{x - Agg(x)\} \tag{5.15}$$

$$r_{x_{i+1}^*, x_i^*} = \max_{x \in [x_i^*, x_{i+1}^*]} \{Agg(x) - x\} \tag{5.16}$$

Proof. We shall give the argument for $r_{x_i^*, x_{i+1}^*}$, the one for $r_{x_{i+1}^*, x_i^*}$ is analogous. We define several groups of players. Let P_d^+ be the set of players who played d in the last period and for whom d is currently a best reply and let P_d^- be the set of players who played d in the last period and for whom m is currently the best reply. Define P_m^+ and P_m^- analogously. Since we consider the resistance between two absorbing states we have for the starting state that $P_d^- = P_m^- = 0$.

Given that $x_i^* < x_{i+1}^*$ we have that in x_{i+1}^* more players play d. Suppose that we construct a path such that only players who currently play m switch to d. Then any such switch is erroneous behavior as long as $x > Agg(x)$. Once $x < Agg(x)$ further transitions have resistance zero. For any given x with $x > Agg(x)$ one needs at least $x - Agg(x)$ errors to enter a region where x_{i+1}^* becomes an attractor. (Note that $x - Agg(x) = x - x_i^*$ for $x_i^* \leq x < x_{i+1}^*$.) Since this must hold for all x with $x > Agg(x)$ we have

$$r_{x_i^*, x_{i+1}^*} = \max_{x \in [x_i^*, x_{i+1}^*]} \{x - Agg(x)\}. \tag{5.17}$$

This is sufficient since after this many trembles there exists a zero resistance path moving to the neighboring absorbing state associated with x_{i+1}^*. ☺

Theorem 36. *Suppose players have* uniform action trembles. *The stochastic potential γ_i of a recurrent class $\boldsymbol{a}_i^* \in \{\boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_l^*\}$ associated with fixpoint $x_i^* \in \{x_1^*, ..., x_l^*\}$ is given by*

$$\gamma_{\boldsymbol{a}_i^*} = \sum_{\alpha=1}^{i-1} r_{x_\alpha, x_{\alpha+1}} + \sum_{\alpha=i+1}^{l} r_{x_\alpha, x_{\alpha-1}}. \tag{5.18}$$

The stochastically stable states are the states associated with fixpoints of Agg that minimize stochastic potential, that is:

$$\lim_{\varepsilon \to 0} \lim_{t \to \infty} \ \bar{a}^t = \{\boldsymbol{a}_i^* : \ \gamma_{\boldsymbol{a}_i^*} = \min_{j=1,\ldots,l} \gamma_{\boldsymbol{a}_j^*}\}. \tag{5.19}$$

For generic games there exists a unique long-term stable state.

Proof. Let $x_1^*, ..., x_l^*$ (in increasing order) be the fixpoints of Agg. In order to pass from one to another one needs to pass through all the fixpoints in between. It follows that the stochastic potential of a fixpoint is given by

$$\gamma_{x_i^*} = \sum_{\alpha=1}^{i-1} r_{x_\alpha, x_{\alpha+1}} + \sum_{\alpha=i+1}^{l} r_{x_\alpha, x_{\alpha-1}}. \tag{5.20}$$

The first summand gives the resistances of passing from the rightmost fixpoint to x_i^*, the second the resistance from passing from the leftmost fixpoint to x_i^*. Now by Young (1993a) we have that the stochastically stable states are precisely those states which have minimal stochastic potential. ☺

5.4.2 Usage

In this section we shall consider the social influence model *Usage*. An active player bases his decision on the cumulative usage (see equation 5.6):

$$s(t) = \frac{\sum_{v=t_0}^{t-1} \sum_{i=1}^{p} \mathbf{1}_{a_i^v = d} \mathbf{1}_i^v}{k} \quad \text{for} \quad t_0 = t - k$$

Theorem 37. *The dynamic has absorbing states if and only if* 0 *and/or* 1 *are fixpoints of Agg or the best response of any player is independent of social influence. In the former case all* $-$ *m and/or all* $-$ *d are the unique absorbing states. If initially players play heterogeneous actions both absorbing states are reached with positive probability. This holds for both k constant and $k = t$.*

Proof. We shall first show that if 0 (or 1) is a fixpoint of *Agg* the corresponding state is absorbing. Suppose that 0 is a fixpoint. Then there exists an observation s^* below which all players want to play m. Suppose that $s(t) < s^*$ in period t. Now independent of who is selected in subsequent periods he plays action m and thus for all $T \geq 0$, $s(t+T) \leq s(t) < s^*$. Hence *all* $-$ m is an absorbing state of the dynamic. If 1 is a fixpoint a similar argument applies.

Now if the best response of any player is independent of social influence there clearly exists an absorbing state, namely the state where every player plays the action he prefers (independent of social influence).

Next suppose that there exists some player for whom social influence may change the best response. Suppose there exists an absorbing state where some players play m and some play d. Suppose a player currently playing m, say i, will be convinced to play d if a high enough proportion of the population plays d (social influence may change his best response). Then, with positive probability enough players who currently wish to play d are selected successively changing the historic action profile such that when i is next selected his best response is d. This shows that there is no mixed-action absorbing state.

By a similar argument one can see that if $all-m$ and $all-d$ are absorbing states of the dynamic either state is attainable when starting with a heterogeneous action profile. Finally note that above arguments hold whether k is fixed or $k = t$. ☺

Perturbed dynamics

We now consider the perturbed *Usage* model with a fixed 'look-back' window.

Theorem 38. *Suppose players have finite 'look-back' of constant size k. Further suppose that the dynamic has no absorbing states. Let $\bar{q}_i = \sum_{j=1}^{i} q_j$. For $i = 1, \ldots, n$ let*

$$\zeta_i := \begin{cases} \frac{1}{\bar{q}_i^{\mu_i}(1-\bar{q}_i)^{1-\mu_i}} \prod_{j=2}^{i} \left(\frac{\bar{q}_{j-1}}{1-\bar{q}_{j-1}}\right)^{\mu_j-\mu_{j-1}} & \textit{if } \mu_i < \bar{q}_i < \mu_{i+1}, \\ 0 & \textit{else.} \end{cases} \tag{5.21}$$

Let $I^* \subseteq \{1, \ldots, n\}$ be such that for $i^* \in I^*$

$$\zeta_{i^*} = \max_{i \in \{1,\ldots,n\}} \zeta_i, \tag{5.22}$$

then

$$\lim_{k\to\infty} \lim_{t\to\infty} \frac{\sum_{v=t-k+1}^{t} \sum_{i=1}^{p} \mathbf{1}_{a_i^v = d} \mathbf{1}_i^v}{t} = \{\bar{q}_{i^*}\}_{i^* \in I^*}. \tag{5.23}$$

$\{\bar{q}_{i^*}\}_{i^* \in I^*}$ is a subset of the fixpoints of Agg. Further, for generic games there exists a unique long-term stable state.

Proof. This result follows from Pinsky (2013, Theorem 4). He studies a random walk on $\mathbb{Z}$ which at each point in time either takes one step to the right or one step to the left. Initially the probability of jumping one step to the right is $\bar{q}_1$ and of jumping to the left $1 - \bar{q}_1$. If the ratio of jumps to the right in the last k moves is greater or equal to μ_i (and smaller than μ_{i+1}) then the probability of jumping to the right is $\bar{q}_i$ and of jumping to the left $1 - \bar{q}_i$. Pinsky studies the 'speed' of the process $\tilde{s} = \lim_{t\to\infty} \frac{X_t}{t}$ where X_t is the position on $\mathbb{Z}$ of the random walk at time t when starting at zero. He finds results according to the theorem above. We shall argue that

$$\frac{\tilde{s}(\cdot) + 1}{2} = \lim_{t\to\infty} \frac{\sum_{v=0}^{t} \sum_{i=1}^{p} \mathbf{1}_{a_i^v = d} \mathbf{1}_i^v}{t}. \tag{5.24}$$

At time t, $\tilde{s}(\cdot) \cdot t$ is the number of times the process stepped to the right minus the number of times the process stepped to the left. Thus in $t - t \cdot \tilde{s}(\cdot)$ steps the process stepped equally often right as left. We note that the sum $\sum_{v=0}^{t} \sum_{i=1}^{p} \mathbf{1}_{a_i^v = d} \mathbf{1}_i^v$ is equivalent to stepping right when Pinsky's process is stepping right and remaining as is when Pinsky's process is stepping left. Thus the transformation follows. ☺

5.4.3 Comparison: Adoption versus Usage

We shall now compare the two different regimes from section 5.4.1 and 5.4.2.[17] We find that, in general, the long-run stable states may differ. Consider the example introduced earlier in this section.

We invite the reader to verify that the fixpoints of Agg are $x_1^* = 2.5\%$, $x_2^* = 16\%$, $x_3^* = 50\%$, $x_4^* = 84\%$, $x_5^* = 97.5\%$. (See Figure 5.2 for an illustration.)

We compute the long-run stable state under *Adoption* according to Theorem 36. One finds the following resistances (up to scaling)

$$r_{x_1^*,x_2^*} = 0.0675,\ r_{x_2^*,x_3^*} = 0.16,\ r_{x_3^*,x_4^*} = 0.18,\ r_{x_4^*,x_5^*} = 0.0675, \tag{5.25}$$

$$r_{x_2^*,x_1^*} = 0.0675,\ r_{x_3^*,x_2^*} = 0.18,\ r_{x_4^*,x_3^*} = 0.16,\ r_{x_5^*,x_4^*} = 0.0675, \tag{5.26}$$

and the stochastic potentials

$$\gamma_{x_1^*} = 0.475,\ \gamma_{x_2^*} = 0.475,\ \gamma_{x_3^*} = 0.455,\ \gamma_{x_4^*} = 0.475,\ \gamma_{x_5^*} = 0.475. \tag{5.27}$$

Thus the (unique) stochastically stable state is x_3^*.

Next, we compute the long-run stable state under *Usage* according to Theorem 38. The rounded results of equation 5.21 are:

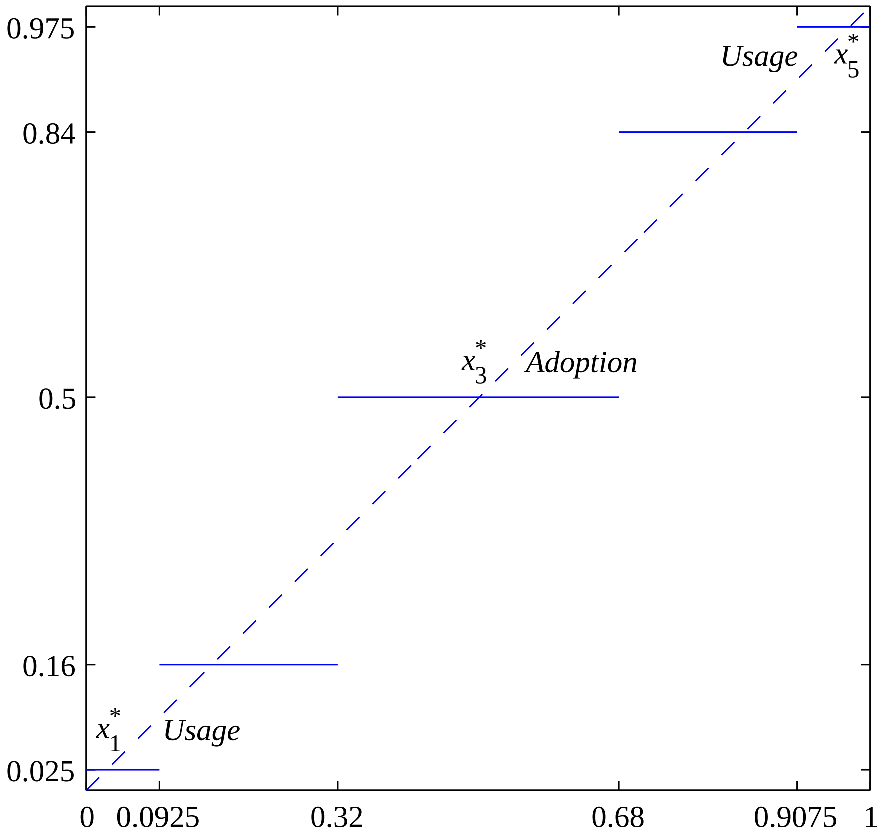
$$x_1^* : 1.026,\ x_2^* : 0.989,\ x_3^* : 0.977,\ x_4^* : 0.989,\ x_5^* : 1.026. \tag{5.28}$$

Thus the long-run stable states are x_1^* and x_5^*.

This shows, by example, that the two different models *Adoption* and *Usage* may yield significantly different outcomes (see Figure 5.3).

[17] Note that the comparison of the unperturbed models immediately follows from Theorems 34 and 37 and reveals that the two models exhibit different behavior.

Figure 5.3: Aggregate dynamic with long-run stable states – x-axis shows μ_i, y-axis shows Agg



Recall our example of mobile phone adoption versus usage of a mobile phone application. Under *Adoption* the shift from one fixpoint to another, say x_i^* to x_{i+1}^*, is governed by the erroneous behavior of players currently *not* playing the innovation (d). In contrast, under *Usage* the shift from one fixpoint to another , say x_i^* to x_{i+1}^*, is governed by higher usage intensity of players currently playing the innovation. More generally, on the one hand, the fluctuation of the number of adopters (away from a fixpoint) does *not* depend on the number of current adopters (given that the fluctuation depends on new adopters). On the other hand, the fluctuation of the cumulative usage depends on the number of current adopters, given that only current adopters have a positive usage.

For two fixpoints, x_i^*, x_j^*, say that x_i^* is *more mixed* (*less mixed*) than x_j^* if $|x_i^* - 0.5| < |x_j^* - 0.5|$ ($|x_i^* - 0.5| > |x_j^* - 0.5|$). Then, on the one hand, the stability of *Adoption* depends on idiosyncratic errors and thus the stability of a fixpoint is independent of whether it is more or less mixed than another fixpoint. On the other hand, the stability under *Usage* is higher for less mixed states, when all else is equal. This is the case since it is less likely for a very small number of users to use a product often enough to 'skew' the observation compared to a more mixed state. In particular suppose we are currently in $x_2^* = 16\%$. In order to reach the basin of attraction of x_3^* the cumulative usage (over the last k periods) needs to be at least 32%. That is, in the last k periods, players for whom d is currently the best response (16% of the population) need to be activated at least $32\% \cdot k$ times. That is, on average such a player needs to be activated at least twice as often as players whose best response is currently m.[18] On the other hand, suppose we are currently in $x_3^* = 50\%$. In order to reach the basin of attraction of x_2^* the cumulative usage (over the last k periods) needs to be at most 32%. That is, in the last k periods, players for whom m is currently the best response (50% of the population) need to be activated at least $(100\% - 32\%) \cdot k$ times. That is, on average such a player needs to be activated at least 1.36 as often as players whose best response is currently d.[19] Since 1.36 is less than 2 it follows that the latter transition is more likely than the former.

5.5 Conclusion

In this chapter we have studied the dynamics of social influence. We considered two different models of social influence. On the one hand, social influence arises from the number of current adopters (of an innovation). On the other hand, social influence arises from the cumulative usage (of an innovation). We

[18]This follows from the simple calculation $32\%/16\% = 2$.

[19]This follows from the simple calculation $(100\% - 32\%)/50\% = 1.36$.

identified the long-run stable states under stochastic dynamics for the two models and found that the outcomes may be very different. The reason being that one model relies on fluctuations by any player while the other model relies on fluctuations by adopters. This suggests that one needs to carefully examine the case of social influence at hand in order to make valuable statements. For example, for technology adoption the driver of social influence in situations of hardware purchases (such as mobile phones) may be *Adoption*. But when considering communication software for mobile phones (such as mobile money) the cumulative usage (*Usage*) is more relevant to understand the process of social influence. Our predictions lend themselves to be tested in an experimental setting or on data which is both left for future research.

Chapter 6

Summary

In order to understand equilibrium play, we need to understand disequilibrium behavior (see, for example, Arrow 1986 and Young 2004). The study of interactive learning dynamics aims to reconcile experimental findings on human behavior with the classical game theoretical analysis which assumes perfect rationality. The latter offers a philosophical foundation of what would be the outcome were everything known to the players and were all players *perfectly* rational. But classic equilibrium analysis does not offer much insight into how people actually behave and agents are typically only *boundedly rational*. Moreover, the environment often features less information than presumed for equilibrium play, and experiments show that players often do not play equilibrium. This is particularly relevant in situations where a large number of players interact and players often meet repeatedly (e.g., online markets, political movements), information is expressly kept private (e.g., negotiations), or situations where players are unaware of the strategic interaction they engage in (e.g., traffic, social norms). Instead players use simple heuristics to choose their strategy based on experimentally tested behavioral rules (see Hoppe 1931, Bush and Mosteller 1951, 1955, Heckhausen 1955 for theoretical models and Tietz and Weber 1972, Tietz 1975, Weber 1976, Tietz et al. 1978, Roth and

Erev 1995, Erev and Roth 1998, Nax et al. 2013a for experiments). By studying these behavioral rules in a dynamic setup we aim to understand if their behaviors converge and if so, to discriminate between different equilibria of the underlying game. This program originates in non-cooperative game theory and has only more recently been studied for cooperative games, sometimes termed as 'naturalizing the social contract' (see, for example, Skyrms 1996, Agastya 1997, 1999, Young 1998, Arnold and Schwalbe 2002, Newton 2010, 2012, Nax 2011, Rozen 2013). Prior to that cooperative game theory considered the question whether cooperative behavior can be explained by non-cooperative play (called the Nash program, see, Nash 1953, Harsanyi 1974, Serrano 2005). In addition, decentralized dynamics have found fruitful applications in computer sciences and control theory, where many machines are programmed according to simple behavioral rules in order to fulfill system-wide tasks.

In this thesis we study decentralized dynamics and learning rules in environments of limited information and bounded rationality. We consider cooperative and non-cooperative games and study whether simple behavioral rules, that are natural consequences of the informational restrictions of the environment, lead to equilibrium play. We then analyze welfare properties of the outcomes, the stability of different equilibria, as well as the time it takes to convergence.

In chapter 2 we provide a completely uncoupled learning rule which converges (in the weak sense of stochastic stability) to the efficient pure Nash equilibria for interdependent, finite games with at least one pure Nash equilibrium. In particular this includes all generic games. This result complements Hart and Mas-Colell (2003) who show that for general games there are no uncoupled dynamics that converge to Nash equilibrium. By weakening the notion of convergence we thus show a positivist counterpart to their result for a particular class of trial and error models. In games that do not have a pure Nash equilibrium, we provide a simple formula that expresses the long-run probability of the various disequilibrium states in terms of two factors: i) the sum of payoffs over all agents, and ii) the maximum payoff gain that results from a unilateral deviation by some agent. This welfare/stability trade-off provides a novel criterion for analyzing the selection of disequilibrium as well as equilibrium states in a large class of games.

On the one hand this result is an existence result stating that there are behaviorally motivated learning rules that lead to efficient equilibria. On the other hand it has already found wide applications in control theory where it is used to program a set of distributed machines to fulfill a system-wide task, such as to maximize the energy production of an off-shore wind farm (Marden et al. 2014) or the energy smart-grid (Cadre and Bedo 2012). Our results and their implications for control theory are discussed in detail in Marden and Shamma (2014) in the 'Handbook of Game Theory' (Young and Zamir 2014). In order to make our design more efficient, future research should try to optimize such decentralized algorithms for specific applications (that is, for specific classes of games) with different informational and computational constraints.

In chapter 3 we study evolutionary dynamics in assignment games (transferable utility, one-to-one matching) and propose a completely uncoupled learning process that selects the core. Adding payoff perturbations we find an equitable selection within the core. Agents encounter each other randomly and make bids for each other and match if profitable. Matches are repeatedly broken and reshuffled. Players on both sides of the market behave in the same way and neither market has an exogenous advantage at any point in time. Equity is favored by the dynamics because it is more stable, not because of any ex ante fairness criterion. This work gives a behavioral justification for the core and its refinement (the least core, Maschler et al. 1979) as solution concepts for decentralized assignment markets. Roth and Vande Vate (1990) showed a somewhat similar result, but in the context of non-transferable utility, one-to-one matching markets (the *marriage market*). Jackson and Watts (2002) study perturbations (matches do not form with small probability) for the marriage market and find that the set of stochastically stable set coincides with the core, a no-selection result in contrast to our least-core selection for the assignment game.

In chapter 4 we study a related dynamic to the one in chapter 3 for the assignment game. Now, there exists an exogenous market sentiment such that sometimes the firms exhibit greater price stickiness than the workers (e.g., in periods of high unemployment), and at other times the reverse holds (e.g., in periods of low unemployment). In addition to convergence to the core we show that this dynamic converges fast, that is, the expected number of steps to convergence scales polynomially in the number of players. Fast converges is an important feature of evolutionary dynamics, since the programmatic aim is to justify equilibrium concepts by behavioral rules. We thus need to ensure that convergence occurs within a reasonable amount of time and is not 'too' dependent on the number of players. Our proof technique develops a new connection between random walks on graphs and matching markets. In future work we would like to develop a deeper understanding of this connection in order to study matching markets and other games on networks.

Finally, in chapter 5 we study social influence in a heterogeneous population and analyze the long-run behavior of these dynamics. Social influence can broadly be separated into two kinds; *Adoption* and *Usage*. For adoption, social influence arises from responding to the number of current adopters. For usage social influence arises from responding to the cumulative usage. We identify the equilibria of the dynamic and show which equilibrium is observed in the long-run. We find that the models exhibit different behavior and hence this differentiation is of importance. We also provide an intuition for the different outcomes. A natural extension of this work is an empirical analysis of the spread of innovations to provide a real-world observation of the importance of this differentiation.

We shall now discuss future routes of research more generally. The questions are intended to be broad and some answers may not be within immediate reach. But we believe that it is important to master the divide between rigorous theory and applicable insights and develop research questions on this ground.

We first recall the motivation of the study of learning in games. The aim is to, in a first step, provide alternative models to the classic full-rationality model which are more likely to describe actual human behavior in a dynamic setting. The main questions are: Do the dynamics converge to a stable state? What are the welfare properties in the long-run? How long does the dynamic take to converge?

Behavioral dynamics – convergence versus chaotic behavior. The literature on convergence of behavioral dynamics in low-information environments is so far split into two camps. On the one hand, there are non-convergence results (see, for example, Hart and Mas-Colell 2003 and Babichenko 2012). On the other hand, there are positivist results showing (weak) convergence to Nash equilibrium (see, for example, Foster and Young 2006, Germano and Lugosi 2007, Young 2009b, and Pradelski and Young 2012). We would like to study under which conditions behavioral rules exhibit convergence and under which conditions chaotic behavior is observed. For example this may depend on the game under consideration, the notion of convergence, or the existence of pure Nash equilibria, etc. In case of chaotic behavior we would like to understand whether the chaotic behavior follows specific patterns. Galla and Farmer (2013) recently studied the emergence of chaotic behavior dependent on the correlations of payoffs for two-player, two-action games for a particular type of reinforcement learning. For pure coordination games their dynamic converges to Nash equilibrium, while for zero sum games it exhibits chaotic behavior. This raises the conjecture that it may be harder to learn mixed equilibrium than pure equilibrium, a potential future route of investigation.

Forward looking, boundedly rational agents. The literature on evolutionary dynamics focuses on myopic agents. That means that players either respond to the current action profile, the recent past, or all the past (see Young 1998 and Young 2004 for a good overview). This may be seen as one of the two extreme points (perfect rationality being the other) of the space of realistic human behavior (compare section 1.2). It is important to understand how the introduction of forward-looking agents would influence the game dynamics. For example one could consider agents responding to the trend rather than the current rate in our model of social influence (chapter 5) or one could revisit the classic adaptive play model (Young 1993a) where players use a myopic best reply to a sample of the past actions. Saez-Marti and Weibull (1999) study a one-step version of forward looking for the latter model. Acemoglu and Jackson (2014) consider a similar type of one-step forward looking for a different dynamic. Going further we would like to investigate how limited forecasting and anticipation influences the game dynamics. Besides being of theoretical interest this would provide greater insights to market behavior as we showed in chapters 3 and 4.

Games with well-ordered strategy sets. In order to study forecasting agents beyond two possible strategies the strategy space needs to be well-ordered, for otherwise forecasting is limited to 'within-strategy' extrapolation.

This points to another important alley of research. In most studies the strategy space is either binary (see, for example, chapter 5) or any finite set without an ordering (see, for example, chapter 2), but a lot of natural heuristics have underlying simple orders, such as amounts, money, height, etc. Understanding behavior under such restricted strategy spaces may allow to explain experiments (where strategies are often well ordered) better and therefore build more realistic models. As a side-effect the study of convergence and rate of convergence can often be reduced to auxiliary variables, thus simplifying the analysis (see, for example, chapter 5).

Medium run behavior, no limits, and heterogeneity. Researchers have so far focused on the long-run behavior of dynamics. As part of this approach it is common to let noise terms vanish (see chapter 1.3). This naturally leaves two questions unanswered. First, what is the medium run behavior of dynamics. Second, interpreting stochastic noise not as erroneous behavior but as actual heterogeneity in the population (for example players may have similar, but different valuations of a good) can we say something about the dynamic without taking noise to zero? These are in itself important questions to advance the study of dynamics but may also have very practical implications. For example, what is the cost in efficiency by implementing a decentralized assignment game (as discussed in chapters 3 and 4) over a centralized assignment game? How will it behave *before* reaching an equilibrium? Interestingly a possible route to understand the medium run of learning dynamics stems from addressing the second question, that is, to avoid taking noise levels to zero. Kreindler and Young (2014) characterize a class of two-action coordination games with non-vanishing noise and nevertheless sharp predictive power. Generalizing such results would be a first step.

Applications. The afore described research agenda can be described as developing theory based on behavioral insights from experiments and the stimulus of other fields such as sociology and psychology. In addition we are interested in identifying currently pressing questions by using the theoretical tools available.

For example, consider the model of social influence discussed in chapter 5. Suppose we want to understand how to disseminate information about Ebola in Africa and have some information or data about previous related attempts. Building a viable model and testing it with available data is the task at hand. Only this will allow us to move from descriptive to prescriptive models. Thus we believe that it is important to study concrete applications where theories will be falsifiable (through data or experiments).

Finally, we would like to mention some future alleys of research concerning mechanism design in control theory. Here one hugely simplifying advantage is that we can program machines in the way we would like them to behave. Our aim is to further deepen our collaborations with the researchers in the respective fields who actually design an off-shore wind farm, say (see, for an overview, Marden and Shamma 2014). This would allow to build models which perform better in terms of stability, efficiency and rate of convergence with the clear aim of immediate applicability.

Bibliography

Acemoglu, D. and M. O. Jackson (2014), "History, expectations, and leadership in the evolution of social norms." *The Review of Economic Studies, forthcoming.*

Ackermann, H., P. Goldberg, V. Mirrokni, H. Roeglin, and B. Voecking (2011), "Uncoordinated two-sided matching markets." *SIAM Journal on Computing*, 40, 92–106.

Agastya, M. (1997), "Adaptive play in multiplayer bargaining situations." *The Review of Economic Studies*, 64, 411–426.

Agastya, M. (1999), "Perturbed adaptive dynamics in coalition form games." *Journal of Economic Theory*, 89, 207–233.

Arieli, I. and Y. Babichenko (2012), "Average testing and Pareto efficiency." *Journal of Economic Theory*, 147, 2376–2398.

Arnold, T. and U. Schwalbe (2002), "Dynamic coalition formation and the core." *Journal of Economic Behavior and Organization*, 49, 363–380.

Arrow, K. J. (1986), "Rationality of self and others in an economic system." *The Journal of Business*, 59, 385–399.

Arthur, W. B. (1991), "Designing economic agents that act like human agents: A behavioral approach to bounded rationality." *American Economic Review*, 81, 353–359.

Asadpour, A. and A. Saberi (2009), "On the inefficiency ratio of stable equilibria in congestion games." *5th Workshop on Internet and Networks Economics*, 5929, 542–552.

Asch, S. E. (1955), "Opinions and social pressure." *Scientific American*, 193, 31–35.

Axelrod, R. (1986), "An evolutionary approach to norms." *The American Political Science Review*, 80, 1095–1111.

Azar, Y., B. Birnbaum, L. E. Celis, N. R. Devanur, and Y. Peres (2009), "Convergence of local dynamics to balanced outcomes in exchange networks." In *FOCS*, 293–302.

Babichenko, Y. (2012), "Completely uncoupled dynamics and nash equilibria." *Games and Economic Behavior*, 76, 1–14.

Babichenko, Y. (2013), "Best-reply dynamics in large binary-choice anonymous games." *Games and Economic Behavior*, 81, 130–144.

Babichenko, Y. (2014), "How long to Pareto efficiency?" *International Journal of Game Theory*, 43, 13–24.

Baddeley, M. (2010), "Herding, social influence and economic decision-making: Socio-psychological and neuroscientific analyses." *Phil. Trans. R. Soc. B*, 365, 281–290.

Baddeley, M., D. Pillas, Y. Christopoulos, W. Schultz, and P. Tobler (2007), "Herding and social pressure in trading tasks: A behavioral analysis." Cambridge Working Papers in Economics No. 0730.

Balinski, M. L. (1965), "Integer programming: Methods, uses, computation." *Management Science*, 12, 253–313.

Balinski, M. L. and D. Gale (1987), "On the core of the assignment game." *Functional Analysis, Optimization and Mathematical Economics*, 1, 274–289.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013), "The diffusion of microfinance." *Science*, 341, 1236498.

Banerjee, A. V. (1992), "A simple model of herd behavior." *Quarterly Journal of Economics*, 107, 797–817.

Bass, F. M. (1969), "A new product growth model for consumer durables." *Management Science*, 15, 215–227.

Bayati, M., C. Borgs, J. Chayes, Y. Kanoria, and A. Montanari (2014), "Bargaining dynamics in exchange networks." Journal of Economic Theory, forthcoming.

Bayati, M., D. Shah, and M. Sharma (2005), *Maximum Weight Matching via Max-Product Belief Propagation*. International Symposium on Information Theory.

Bayati, M., D. Shah, and M. Sharma (2008), "Max-product for maximum weight matching: convergence, correctness, and lp duality." *IEEE Transactions on information theory*, 54, 1241–1251.

Beggs, A.W. (2005), "On the convergence of reinforcement learning." *Journal of Economic Theory*, 122, 1–36.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992), "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy*, 100, 992–1026.

Biró, P., M. Bomhoff, P. A. Golovach, W. Kern, and D. Paulusma (2012), "Solutions for the stable roommates problem with payments." *Lecture Notes in Computer Science*, 7551, 69–80.

Blume, L. E. (1993), "The statistical mechanics of strategic interaction." *Games and Economic Behavior*, 5, 387–424.

Blume, L. E. (1995), "The statistical mechanics of best-response strategy revision." *Games and Economic Behavior*, 11, 111–145.

Blume, L. E. (2003), "How noise matters." *Games and Economic Behavior*, 44, 251–271.

Bongaarts, J. and S. C. Watkins (1996), "Social interactions and contemporary fertility transitions." *Population and Development Review*, 22, 639–682.

Bouchaud, J.-P. (2013), "Crises and collective socio-economic phenomena. cartoon models and challenges." *Journal of Statistical Physics*, 151, 567–606.

Bowling, M. and M. Veloso (2002), "Multi-agent learning with a variable learning rate." *Artificial Intelligence*, 136, 215–250.

Bramoulle, Y. (2007), "Anti-coordination and social interactions." *Games and Economic Behavior*, 58, 30–49.

Bush, R. and F. Mosteller (1951), "A mathematical model for simple learning." *Psychological Review*, 58, 313–323.

Bush, R. and F. Mosteller (1955), *Stochastic Models of Learning*. Wiley.

Cabinet Office (2012), *Applying behavioural insights to reduce fraud, error and debt*. Behavioural Insights Team UK.

Cadre, H. and S. Bedo (2012), "Distributed learning in hierarchical networks." *6th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS)*, 1, 188–197.

Celis, L. E., N. R. Devanur, and Y. Peres (2010), "Local dynamics in bargaining networks via random-turn games." *WINE 2010, LNCS, Springer*, 6484, 133–144.

Centola, D. and M. Macy (2007), "Complex contagions and the weakness of long ties." *American Journal of Sociology*, 113, 702–734.

Chen, B., S. Fujishige, and Z. Yang (2011), "Decentralized market processes to stable job matchings with competitive salaries." KIER Working Papers 749, Kyoto University.

Chung, K.-S. (2000), "On the existence of stable roommate matching." *Games and Economic Behavior*, 33, 206–230.

Cont, R. and J.-P. Bouchaud (2000), "Herd behavior and aggregate fluctuations in financial markets." *Macroeconomic Dynamics*, 4, 170–196.

Crawford, V. P. and E. M. Knoer (1981), "Job matching with heterogeneous firms and workers." *Econometrica*, 49, 437–540.

Cukierman, A. (1991), "Asymmetric information and the electoral momentum of public opinion polls." *Public Choice*, 70, 181–213.

Davis, M. and M. Maschler (1965), "The kernel of a cooperative game." *Naval Research Logistic Quarterly*, 12, 223–259.

de Silva, H., D. Ratnadiwakara, and A. Zainudeen (2011), "Social influence in mobile phone adoption: Evidence from the bottom of the pyramid in emerging asia." *Information Technologies & International Development*, 7, 1–18.

Demange, G. and D. Gale (1985), "The strategy of two-sided matching markets." *Econometrica*, 53, 873–988.

Demange, G., D. Gale, and M. Sotomayor (1986), "Multi-item auctions." *Journal of Political Economy*, 94, 863–872.

Devenow, A. and I. Welch (1996), "Rational herding in financial economics." *European Economic Review*, 40, 603–615.

Diamantoudi, E., L. Xue, and E. Miyagawa (2004), "Random paths to stability in the roommate problem." *Games and Economic Behavior*, 48, 18–28.

Dodds, P. S. and D. J. Watts (2004), "Universal behavior in a generalized model of contagion." *Physical Review Letters*, 92, 218701.

Dodds, P. S. and D. J. Watts (2005), "A generalized model of social and biological contagion." *Journal of Theoretical Biology*, 232, 587–604.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2012), "The intergenerational transmission of risk and trust attitudes." *The Review of Economic Studies*, 79, 645–677.

Driessen, T. S. H. (1998), "A note on the inclusion of the kernel in the core for the bilateral assignment game." *International Journal of Game Theory*, 27, 301–303.

Driessen, T. S. H. (1999), "Pairwise-bargained consistency and game theory: the case of a two-sided firm." *Fields Institute Communications*, 23, 65–82.

Edmonds, J. (1965), "Paths, trees, and flowers." *Canadian Journal of Mathematics*, 17, 49–467.

Edmonds, J. and R. M. Karp (1972), "Theoretical improvements in algorithmic efficiency for network flow problems." *Journal of the ACM*, 19, 248–264.

Ellison, G. (1993), "Learning, local interaction, and coordination." *Econometrica*, 61, 1047–1071.

Ellison, G. and D. Fudenberg (1995), "Word of mouth, communication and social learning." *Quarterly Journal of Economics*, 110, 93–125.

Erev, I. and A. E. Roth (1998), "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria." *American Economic Review*, 88, 848–881.

Estes, W. (1950), "Towards a statistical theory of learning." *Psychological Review*, 57, 94–107.

Feller, W. (1950), *An introduction to probability theory and ist applications.* John Wiley.

Foster, D. and H. P. Young (1990), "Stochastic evolutionary game dynamics." *Theoretical Population Biology*, 38, 219–232.

Foster, D. P. and R. Vohra (1997), "Calibrated learning and correlated equilibrium." *Games and Economic Behavior*, 21, 40–55.

Foster, D. P. and H. P. Young (2003), "Learning, hypothesis testing, and nash equilibrium." *Games and Economic Behavior*, 45, 73–96.

Foster, D. P. and H. P. Young (2006), "Regret testing: Learning to play Nash equilibrium without knowing you have an opponent." *Theoretical Economics*, 1, 341–367.

Freud, S. (1921a), *Group psychology and the analysis of the ego (translation by J. Strachey).* The International Psycho-Analytical Library.

Freud, S. (1921b), *Massenpsychologie und Ich-Analyse.* Internationaler Psychoanalytischer Verlag.

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

Gale, D. and L. S. Shapley (1962), "College admissions and stability of marriage." *American Mathematical Monthly*, 69, 9–15.

Galla, T. and J. D. Farmer (2013), "Complex dynamics in learning complicated games." *PNAS*, 110, 1232–1236.

Georgakopoulos, A. and S. Wagner (2013), "Hitting times, cover costs, and the Wiener index of a tree." Working Paper.

Germano, F. and G. Lugosi (2007), "Global Nash convergence of Foster and Young's regret testing." *Games and Economic Behavior*, 60, 135–154.

Gillies, D. B. (1959), "Solutions to general non-zero-sum games." *H. Kuhn and A. Tucker (Eds.) Contributions to the theory of games*, 4, 47–85.

Golub, B. and M. O. Jackson (2010), "Naive learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics*, 2, 112–149.

Granovetter, M. (1978), "Threshold models of collective behavior." *The American Journal of Sociology*, 83, 1420–1443.

Hamilton, W. D. (1971), "Geometry for the selfish herd." *Journal of Theoretical Biology*, 31, 295–311.

Harsanyi, J. C. (1974), "An equilibrium-point interpretation of stable sets and a proposed alternative definition." *Management Science*, 20, 1472–1495.

Hart, S. and Y. Mansour (2010), "How long to equilibrium? The communication complexity of uncoupled equilibrium procedures." *Games and Economic Behavior*, 69, 107–126.

Hart, S. and A. Mas-Colell (2000), "A simple adaptive procedure leading to correlated equilibrium." *Econometrica*, 68, 1127–1150.

Hart, S. and A. Mas-Colell (2003), "Uncoupled dynamics do not lead to Nash equilibrium." *American Economic Review*, 93, 1830–1836.

Hart, S. and A. Mas-Colell (2006), "Stochastic uncoupled dynamics and Nash equilibrium." *Games and Economic Behavior*, 57, 286–303.

Hart, S. and A. Mas-Colell (2013), *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*. World Scientific Publishing Company.

Heckhausen, H. (1955), "Motivationsanalyse der Anspruchsniveau-Setzung." *Psychologische Forschung*, 25, 118–154.

Herrnstein, R. J. (1961), "Relative and absolute strength of response as a function of frequency of reinforcement." *Journal of Experimental Analysis of Behavior*, 4, 267–272.

Hong, H., J. D. Kubik, and J. C. Stein (2004), "Social interaction and stock-market participation." *The Journal of Finance*, 59, 137–163.

Hoppe, F. (1931), "Erfolg und Mißerfolg." *Psychologische Forschung*, 14, 1–62.

Horst, U. and J. A. Scheinkman (2006), "Equilibria in systems of social interactions." *Journal of Economic Theory*, 130, 44–77.

Houston, A. I., A. Kacelnik, and J. M. McNamara (1982), *Some learning rules for acquiring information*. Functional Ontogeny, D. J. McFarland, ed., New York: Pitman.

Hu, Y., B. Skyrms, and P. Tarres (2011), "Reinforcement learning in signaling game." ArXiv:1103.5818v1 [math.PR] 30 Mar 2011.

Huberman, G. (1980), "The nucleolus and the essential coalitions." *Analysis and Optimization of Systems, Lecture Notes in Control and Information Systems*, 28, 417–422.

Inarra, E., C. Larrea, and E. Molis (2008), "Random paths to P-stability in the roommate problem." *International Journal of Game Theory*, 36, 461–471.

Inarra, E., C. Larrea, and E. Molis (2013), “Absorbing sets in roommate problem.” *Games and Economic Behavior*, 81, 165–178.

Jackson, M. O. and A. Watts (2002), “The evolution of social and economic networks.” *Journal of Economic Theory*, 106, 265–295.

Kandori, M., G. J. Mailath, and R. Rob (1993), “Learning, mutation, and long run equilibria in games.” *Econometrica*, 61, 29–56.

Kanoria, Y., M. Bayati, C. Borgs, J. Chayes, and A. Montanari (2011), “Fast convergence of natural bargaining dynamics in exchange networks.” In *Proc. ACM-SIAM Symp. On Discrete Algorithms*, 1518–1537.

Karandikar, R., D. Mookherjee, D. Ray, and F. Vega-Redondo (1998), “Evolving aspirations and cooperation.” *Journal of Economic Theory*, 80, 292–331.

Kelso, A. S. and V. P. Crawford (1982), “Job matching, coalition formation, and gross substitutes.” *Econometrica*, 50, 1483–1504.

Kermack, W. O. and A. G. McKendrick (1927), “A contribution to the mathematical theory of epidemics.” *Proceedings of the Royal Society*, 115, 700–721.

Keynes, J. M. (1930), *A treatise on money.* Macmillan.

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money.* Macmillan.

Kim, J. H. and J. Pearl (1983), “A computational model for causal and diagnostic reasoning in inference systems.” *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 83, 190–193.

Kiringai, J. and W. Fengler (2010), “Kenya at the tipping point? with a special focus on the ICT revolution and mobile money: Kenya economic report.” Kenya economic update ; edition no. 3. Washington, DC: World Bank.

Klaus, B. and F. Klijn (2007), “Paths to stability for matching markets with couples.” *Games and Economic Behavior*, 58, 154–171.

Klaus, B., F. Klijn, and M. Walzl (2010), “Stochastic stability for roommates markets.” *Journal of Economic Theory*, 145, 2218–2240.

Klaus, B. and F. Payot (2013), “Paths to stability in the assignment problem.” Working paper.

Kleinberg, J. and E. Tardos (2008), “Balanced outcomes in social exchange networks.” In *43rd ACM Symposium on Theory of Computing.*

Kojima, F. and M. U. Ünver (2008), “Random paths to pairwise stability in many-to-many matching problems: a study on market equilibrium.” *International Journal of Game Theory*, 36, 473–488.

Kreindler, G. E. and H. P. Young (2013), “Fast convergence in evolutionary equilibrium selection.” *Games and Economic Behavior*, 80, 39–67.

Kreindler, G. E. and H. P. Young (2014), "Rapid innovation diffusion in social networks." *PNAS*, 111, 10881–10888.

Kuhn, H. W. (1955), "The Hungarian Method for the assignment problem." *Naval Research Logistic Quarterly*, 2, 83–97.

Le Bon, G. (1895), *Psychologie des foules*.

Llerena, F. and M. Nunez (2011), "A geometric characterization of the nucleolus of the assignment game." *Economics Bulletin*, 31, 3275–3285.

Llerena, F., M. Nunez, and C. Rafels (2015), "An axiomatization of the nucleolus of assignment markets." *International Journal of Game Theory*, 44, 1–15.

Lohmann, S. (1994), "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989-91." *World Politics*, 47, 42–101.

Lopez-Pintado, D. and D. J. Watts (2008), "Social influence, binary decisions and collective dynamics." *Rationality and Society*, 20, 399–443.

Macy, M. W. (1991), "Chains of cooperation: Threshold effects in collective action." *American Sociological Review*, 56, 730–747.

Mannor, S. and J. S. Shamma (2007), "Multi-agent learning for engineers." *Artificial Intelligence*, 171, 417–422.

Marden, J. R., G. Arslan, and J. S. Shamma (2009a), *Cooperative control and potential games*. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics.

Marden, J. R. and J. S. Shamma (2012), "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation." *Games and Economic Behavior*, 75, 788–808.

Marden, J. R. and J. S. Shamma (2014), "Game Theory and Distributed Control." *Handbook of Game Theory, edited by Peyton Young and Shmuel Zamir, Elsevier Science*, 4, 861–900.

Marden, J. R., H. P. Young, G. Arslan, and J. Shamma (2009b), "Payoff-based dynamics for multi-player weakly acyclic games." *SIAM Journal on Control and Optimization*, 48, 373–396.

Marden, J. R., H. P. Young, and L. Y. Pao (2014), "Achieving Pareto optimality through distributed learning." *SIAM Journal on Control and Optimization*, 52, 2753–2770.

Maschler, M., B. Peleg, and L. S. Shapley (1979), "Geometric properties of the kernel, nucleolus, and related solution concepts." *Mathematics of Operations Research*, 4, 303–338.

Maynard Smith, J. and G. R. Price (1973), "The logic of animal conflict." *Nature*, 246, 15–18.

McKelvey, R. D. and T. R. Palfrey (1995), "Quantal response equilibria for normal form games." *Games and Economic Behavior*, 10, 6–38.

Meade, N. and T. Islam (2006), "Modelling and forecasting the diffusion of innovation - a 25-year review." *International Journal of Forecasting*, 22, 519–545.

Montaneri, A. and A. Saberi (2010), "The spread of innovations in social networks." *PNAS*, 107, 20196–20201.

Montgomery, M. R. and J. B. Casterline (1996), "Social learning, social influence, and new models of fertility." *Population and Development Review*, 22, 151–175.

Morris, S. (2000), "Contagion." *The Review of Economic Studies*, 67, 57–78.

Motro, U. and A. Shmida (1995), "Near-far search: an evolutionarily stable foraging strategy." *Journal of Theoretical Biology*, 173, 15–22.

Moussaid, M., J. E. Kaemmer, P. P. Analytis, and H. Neth (2013), "Social influence and the collective dynamics of opinion formation." *PloS one*, 8, e78433.

Nagel, R. (1995), "Unraveling in guessing games: An experimental study." *American Economic Review*, 85, 1313–1326.

Nash, J. (1950a), "The bargaining problem." *Econometrica*, 18, 155–162.

Nash, J. (1950b), *Non-cooperative games*. Ph.D. thesis, Princeton University.

Nash, J. (1953), "Two-person cooperative games." *Econometrica*, 21, 128–140.

Nax, H. H. (2011), *Evolutionary Cooperative Games*. Ph.D. thesis, University of Oxford.

Nax, H. H., M. N. Burton-Chellew, S. A. West, and H. P. Young (2013a), "Learning in a black box: The case of voluntary contribution games." Department of Economics Discussion Paper Series 653, University of Oxford.

Nax, H. H. and B. S. R. Pradelski (2014), "Evolutionary dynamics and equitable core selection in assignment games." International Journal of Game Theory, forthcoming.

Nax, H. H., B. S. R. Pradelski, and H. P. Young (2013b), "Decentralized dynamics to optimal and stable states in the assignment game." In *Proceedings of the 52nd IEEE Conference on Decision and Control, December 10-13, 2013. Florence, Italy*, 2391–2397.

Nevo, I. and I. Erev (2012), "On surprise, change, and the effect of recent outcomes." *Frontiers in Psychology*, 3, 1–9.

Newton, J. (2010), *Non-cooperative convergence to the core in Nash demand games without random errors or convexity assumptions*. Ph.D. thesis, University of Cambridge.

Newton, J. (2012), "Recontracting and stochastic stability in cooperative games." *Journal of Economic Theory*, 147, 364–381.

Newton, J. and R. Sawa (2015), "A one-shot deviation principle for stability in matching problems." *Journal of Economic Theory*, 157, 1–27.

Norris, J. R. (1997), *Markov Chains*. Cambridge University Press.

Nunez, M. (2004), "A note on the nucleolus and the kernel of the assignment game." *International Journal of Game Theory*, 33, 55–65.

Papadimitriou, C. (2001), "Algorithms, games and the internet." In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing*, 749–753.

Pearl, J. (1982), "Reverend bayes on inference engines: A distributed hierarchical approach." *Proceedings of the Second National Conference on Artificial Intelligence*, 82, 133–136.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann.

Peleg, B. and P. Sudholter (2007), *Introduction to the Theory of Cooperative games*. Springer-Verlag.

Pinsky, R. G. (2013), "The speed of a random walk excited by its recent history." Working Paper, arXiv: 1305.7242.

Pradelski, B. S. R. and H. P. Young (2012), "Learning efficient Nash equilibria in distributed systems." *Games and Economic Behavior*, 75, 882–897.

Richardson, T. and R. Urbanke (2001), "The capacity of low-density parity check codes under message-passing decoding." *IEEE Transactions on information theory*, 47, 599–618.

Rochford, S. C. (1984), "Symmetrically pairwise-bargained allocations in an assignment market." *Journal of Economic Theory*, 34, 262–281.

Rogers, E. M. (1962), *Diffusion of Innovations*. The Free Press.

Roth, A. E. and I. Erev (1995), "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term." *Games and Economic Behavior*, 8, 164–212.

Roth, A. E. and M. Sotomayor (1990), *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. Cambridge University Press.

Roth, A. E. and M. Sotomayor (1992), "Two-sided matching." *Handbook of Game Theory with Economic Applications*, 1, 485–541.

Roth, A. E. and H. Vande Vate (1990), "Random paths to stability in two-sided matching." *Econometrica*, 58, 1475–1480.

Roughgarden, T. (2005), *Selfish routing and the price o anarchy*. The MIT Press.

Rozen, K. (2013), "Conflict leads to cooperation in Nash Bargaining." *Journal of Economic Behavior and Organization*, 87, 35–42.

Saez-Marti, M. and J. W. Weibull (1999), "Clever agents in Young's evolutionary bargaining model." *Journal of Economic Theory*, 86, 268–279.

Salez, J. and D. Shah (2009), "Belief propagation: an asymptotically optimal algorithm for the random assignment problem." *Mathematics of Operations Research*, 34, 468–480.

Salganik, M. J., P. S. Dodds, and D. J. Watts (2006), "Experimental study of inequality and unpredictability in an artificial cultural market." *Science*, 311, 854–856.

Sandholm, W. H. (2002), "Evolutionary implementation and congestion pricing." *The Review of Economic Studies*, 69, 667–689.

Sandholm, W. H. (2010), *Population games and evolutionary dynamics*. The MIT Press.

Sauermann, H. and R. Selten (1962), "Anspruchsanpassungstheorie der Unternehmung." *Zeitschrift fuer die gesamte Staatswissenschaft*, 118, 577–597.

Sawa, R. (2014), "Coalitional stochastic stability in games, networks and markets." *Games and Economic Behavior*, 88, 90–111.

Scharfstein, D. S. and J. C. Stein (1990), "Herd behavior and investment." *American Economic Review*, 80, 465–479.

Scheinkman, J. A. (2008), *Social interactions*. The New Palgrave Dictionary of Economics, 2nd edition, S. Durlauf and L. Blume (eds.), Palgrave Macmillan.

Schelling, T. C. (1971), "Dynamic models of segregation." *Journal of Mathematical Sociology*, 1, 143–186.

Schelling, T. C. (1978), *Micromotives and Macrobehavior*. Norton & Company.

Schmeidler, D. (1969), "The nucleolus of a characteristic function game." *SIAM Journal of Applied Mathematics*, 17, 1163–1170.

Scholz, R. W., A. Fleischer, and A. Bentrup (1983), "Aspiration forming and predictions based on aspiration levels compared between professional and non-professional bargainers." *R. Tietz (Ed.), Aspiration levels in bargaining and economic decision making, Lecture Notes in Economics and Mathematical Systems*, 213, 104–121.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007), "The constructive, destructive and reconstructive power of social norms." *Psychological Science*, 18, 429–434.

Selten, R. (1998), "Aspiration adaptation theory." *Journal of Mathematical Psychology*, 42, 191–214.

Selten, R. and R. Stoecker (1986), "End behavior in sequences of finite prisoner's dilemma supergames: A learning theory approach." *Journal of Economic Behavior and Organization*, 7, 47–70.

Serrano, R. (2005), "Fifty years of the nash program, 1953-2013." *Investigaciones Economicas, Fundacin SEPI*, 29, 219–258.

Shah, D. and J. Shin (2010), "Dynamics in congestion games." In *SIGMETRICS '10 Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, 107–118.

Shapley, L. S. (1953), "A value for n-person games." *Contributions to the Theory of Games, volume II, by H.W. Kuhn and A.W. Tucker, editors. Annals of Mathematical Studies*, 28, 307–317.

Shapley, L. S. and M. Shubik (1963), "The core of an economy with nonconvex preferences." *The Rand Corporation*, 3518.

Shapley, L. S. and M. Shubik (1966), "Quasi-cores in a monetary economy with nonconvex preferences." *Econometrica*, 34.

Shapley, L. S. and M. Shubik (1972), "The assignment game 1: The core." *International Journal of Game Theory*, 1, 111–130.

Shiller, R. J. (1984), "Stock prices and social dynamics." *Brookings Papers on Economic Activity*, 2, 457–498.

Shiller, R. J. (2000), *Irrational exuberance*. Princeton University Press.

Skyrms, B. (1996), *Evolution of the social contract*. Cambridge University Press.

Solymosi, T. and T. E. S. Raghavan (1994), "An algorithm for finding the nucleolus of assignment games." *International Journal of Game Theory*, 23, 119–143.

Sotomayor, M. (2003), "Some further remarks on the core structure of the assignment game." *Mathematical Social Sciences*, 46, 261–265.

Stahl, D. O. and P. W. Wilson (1995), "On players' models of other players: Theory and experimental evidence." *Games and Economic Behavior*, 10, 218–254.

Sturgeon, S. (2009), "Belief, reason and logic." *Royal Institute of Philosophy Supplement*, 64, 89–100.

Teng, J. T. C., V. Grover, and W. Guettler (2002), "Information technology innovations: General diffusion patterns and its relationships to innovation characteristics." *IEEE Transactions on Engineering Management*, 49, 13–27.

Tetali, P. and P. Winkler (1991), "On a random walk problem arising in self-stabilizing token management." In *Proceedings of the tenth annual ACM* *symposium on Principles of distributed computing*, PODC '91, 273–280, ACM, New York, NY, USA, URL `10.1145/112600.112623`.

Thorndike, E. (1898), "Animal intelligence: An experimental study of the associative processes in animals." *Psychological Review*, 8, 1874–1949.

Thuijsman, F., B. Peleg, M. Amitai, and A. Shmida (1995), "Automata, matching, and foraging behavior in bees." *Journal of Theoretical Biology*, 175, 305–316.

Tietz, A. and H. J. Weber (1972), "On the nature of the bargaining process in the Kresko-game." *H. Sauermann (Ed.) Contribution to experimental economics*, 7, 305–334.

Tietz, R. (1975), "An experimental analysis of wage bargaining." *Zeitschrift fuer die gesamte Staatswissenschaft*, 131, 44–91.

Tietz, R., H. Weber, U. Vidmajer, and C. Wentzel (1978), "On aspiration forming behavior in repetitive negotiations." *Contributions to experimental economics, H. Sauermann (ed.)*, 7, 88–102.

Trotter, W. (1916), *Instincts of the Herd in War and Peace*.

Valente, T. W. (1996), "Social network thresholds in the diffusion of innovations." *Social Networks*, 18, 69–89.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton University Press.

Weber, H. (1976), "On the theory of adaptation of aspiration levels in a bilateral decision setting." *Zeitschrift fuer die gesamte Staatswissenschaft*, 132, 582–591.

Weibull, J. (1995), *Evolutionary Game Theory*. The MIT Press.

Weibull, J. (1998), "What have we learned from evolutionary game theory so far?" Working Paper.

Welch, I. (1992), "Sequential sales, learning, and cascades." *The Journal of Finance*, 47, 695–732.

Young, H. P. (1993a), "The evolution of conventions." *Econometrica*, 61, 57–84.

Young, H. P. (1993b), "An evolutionary model of bargaining." *Journal of Economic Theory*, 59, 145–168.

Young, H. P. (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

Young, H. P. (2004), *Strategic learning and its limits*. Oxford University Press.

Young, H. P. (2009a), "Innovation diffusion in heterogeneous populations: contagion, social influence, and social learning." *American Economic Review*, 99, 1899–1924.

Young, H. P. (2009b), “Learning by trial and error.” *Games and Economic Behavior*, 65, 626–643.

Young, H. P. (2011), “The dynamics of social innovation.” *PNAS*, 108, 21285–21291.

Young, H. P. and S. Zamir (2014), *Handbook of Game Theory, Volume 4*. North Holland.

Yu, C.-S. (2012), “Factors affecting individuals to adopt mobile banking: Empirical evidence from the utaut model.” *Journal of Electronic Commerce Research*, 13, 104–121.