Evolution and Learning in Games

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.
EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization
The research activities are organized in twenty Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT	
Management and Organisation; (A)	Prof Sven-Erik Sjöstrand
Center for Ethics and Economics; (CEE)	Adj Prof Hans de Geer
Center for Entrepreneurship and Business Creation; (E)	Prof Carin Holmquist
Public Management; (F)	Prof Nils Brunsson
Information Management; (I)	Prof Mats Lundeberg
Center for People and Organization; (PMO)	Prof Jan Löwstedt
Center for Innovation and Operations Management; (T)	Prof Christer Karlsson
ECONOMIC PSYCHOLOGY	
Center for Risk Research; (CFR)	Prof Lennart Sjöberg
Economic Psychology; (P)	Prof Lennart Sjöberg
MARKETING	
Center for Information and Communication Research; (CIC)	Adj Prof Bertil Thorngren
Center for Consumer Marketing; (CCM)	Acting Prof Magnus Söderlund
Marketing, Distribution and Industrial Dynamics; (D)	Prof Lars-Gunnar Mattsson
ACCOUNTING, CONTROL AND CORPORATE FINANCE	
Accounting and Managerial Finance; (B)	Prof Lars Östman
Managerial Economics; (C)	Prof Peter Jennergren
FINANCE	
Finance; (FI)	Prof Clas Bergström
ECONOMICS	
Center for Health Economics; (CHE)	Prof Bengt Jönsson
International Economics and Geography; (IEG)	Prof Mats Lundahl
Economics; (S)	Prof Lars Bergman
ECONOMICS STATISTICS	
Economic Statistics; (ES)	Prof Anders Westlund
LAW	
Law; (RV)	Prof Erik Nerep

Chairman of the Board: Prof Håkan Lindgren. *Director:* Associate Prof Bo Sellstedt.

Adress
EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Evolution and Learning in Games

Jens Josephson

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics, 2001

© EFI and the author
ISBN nr 91-7258-587-0

Keywords:
Bounded rationality, Evolutionary game theory, Learning rules,
Markov chain, Stochastic stability, Evolutionary stability,
Imitation, Better replies, Heterogeneous populations,
Belief learning, Reinforcement learning

Cover photo:
"Det sjunde inseglet" ("The Seventh Seal")
© 1957 AB Svensk Filmindustri
Photographer: Louis Huch

Printed by Elanders Gotab, Stockholm, 2001

Distributed by:
EFI, The Economic Research Institute
Stockholm School of Economics
P. O. Box 6501, SE 1-113 83 Stockholm, Sweden
www.hhs.se/efi

Acknowledgements

First of all, I wish to express my sincere gratitude to my supervisor, Jörgen W. Weibull, who influenced the contents of this thesis through numerous inspiring lectures and discussions, and encouraged me when I felt there were better things to do than economic research.

I also want to thank all other people who have commented on essays in this thesis, such as Ken Binmore, Martin Dufwenberg, Ariel Rubinstein, Maria Sáez Martí, William Sandholm, Philippe Solal, and seminar participants at the Stockholm School of Economics, Cornell University, and various conferences. Special thanks to Bo Becker, who was instrumental in my decision to return to academia, and Alexander Matros, coauthor of one of the essays in this thesis.

I am grateful to Larry Blume and David Easley for allowing me to visit the Department of Economics of Cornell University during the spring term of 2001.

I am indebted to Stefan Engström and Sune Karlsson for answering my questions regarding statistical software, and to Pirjo Furtenbach, Ritva Kiviharju, Kerstin Niklasson, and Ingrid Nilsson, who all helped me with administrative and other matters during these years of study and writing.

Financial support from Jan Wallander and Tom Hedelius Stiftelse, and Bankforskningsinstitutet is gratefully acknowledged.

Last, but not least, I wish to thank Isabel and my family, who put up with me throughout these years.

All remaining errors are, of course, my own.

Stockholm, November 2001

Jens Josephson

Contents

[1] Coauthored with Alexander Matros.

"It is not the strongest of the species that survive, nor the most intelligent, but the one most responsive to change."
– Charles Darwin

Introduction

This thesis contains four separate essays on evolution and learning in games. As an introduction, I will first provide a short description of the field of *evolutionary game theory*, and a brief summary of the different essays.[2]

0.1 Evolutionary Game Theory

Game theory is the theory of strategic interaction between decision makers.[3] The reason for its, somewhat misleading, name (and for the cover photo of this thesis) is that some of its earliest applications were studies of parlor games, such as poker and chess. However, game theory has a much broader scope. It provides a general mathematical technique for analyzing situations where two parties or more make decisions with an impact on each other's welfare. Game theory can be applied to all social sciences, but so far its most widespread use has been within the field of economics. This is not surprising, given that economic settings almost always contains an element of strategic interaction; just think firms competing in a market, employers and employees negotiating within a firm, auctions, or different shareholders trying to gain control over a company.

Game theory is usually divided into two branches: *cooperative* and *non-cooperative* game theory. Within cooperative game theory, the joint actions of groups of decision-makers are primitives. Within noncooperative game theory, which is the focus of this thesis, the individual actions of the decision-makers are primitives.

More specifically, the central object in noncooperative game theory is that of a *game*, which consists of three elements. First, a set of interacting parties, or *players*, acting in their own perceived interest. Second, a set of *strategies* for each player. Strategies are plans that prescribe an action for any possible

[2] For a more detailed introduction to evolutionary game theory, see e.g. Samuelsson (1997), Vega-Redondo (1996), Weibull (1995), Weibull (1997), and Young (1998).

[3] For an introduction to game theory, see e.g. Myerson (1991), Fudenberg and Tirole (1991), Ritzberger (2001), and Vega-Redondo (2001).

contingency. A vector consisting of one strategy for each player is called a *strategy profile.* Third, a set of *preferences* over all possible strategy profiles, one preference ordering for each player. Preferences are generally represented by a *payoff function.*

The most important concept within noncooperative game theory is that of *Nash equilibrium*, due to Nash (1950). A Nash equilibrium is a strategy profile, such that each player maximizes his own payoff given his belief about the other players' choice of strategies, and such that all players' beliefs are correct.

Consider the following example. Two firms producing mobile phones simultaneously and independently need to decide which one two standards to adopt. It is in both firms' interest to choose the same standard as the other firm. If they both choose the old standard A, they both expect to earn a profit of 3 units. If they both invest in the new standard B, they both expect to earn a profit of 4 units. If, on the other hand, one of the firms invests in standard A, and the other firm in standard B, the firm investing in A expects to earn 2 units and the other firm 0 units. This situation can be modeled as a *Coordination Game* where the two firms are players. Player 1 has the two *pure* (i.e. non-randomizing) strategies, A and B, and Player 2 the pure strategies a and b. Hence, there are four pure-strategy profiles, (A, a), (A, b), (B, a), and (B, b). The payoffs to each strategy profile are illustrated in Figure 0.1. Player 1's payoff is indicated by the the left-most figure in the corresponding cell, and Player 2's payoff by the right-most figure. This game has two Nash equilibria in pure strategies; (A, a), the *risk-dominant* equilibrium, and (B, b) the *Pareto-dominant* equilibrium. There is also a third Nash equilibrium in *mixed* strategies (i.e. where each player assigns a probability to each strategy and randomizes), where Player 1 (2) plays strategy A (a) with probability $\frac{2}{5}$ and strategy B (b) with probability $\frac{3}{5}$. In all these three equilibria, each player is maximizing his expected payoff given the other player's strategy.

	a	b
A	3, 3	2, 0
B	0, 2	4, 4

FIGURE 0.1

Traditional noncooperative game theory assumes that a game is played only once, and that a Nash equilibrium arises through a mental process of deductive reasoning by each player. This process does not only require that the players are fully rational, but also that the rules of the game, the rationality of the players, the players' payoff functions, and the predicted equilibrium are all common knowledge, i.e. all players know these things, know that all players know these things, etc. ad infinitum.[4] These assumptions are not realistic in almost any economic settings. One way of getting around this problem is to dispense with the assumption that the players are fully rational and instead assume that they are boundedly rational, but play the game recurrently and adapt their play in light of their experience, and to ask if they in the long run will behave *as if* they were rational, perhaps even as if rationality were common knowledge etc. Under this hypothesis, rational and equilibrium behavior are not a result of introspection, but the possible consequences of an evolutionary or learning process. In the non-strategic context of competitive markets, this Darwinian idea has been attributed to Friedman (1953) and Alchian (1950), but within the strategic context of game theory, Nash (1950) provided a similar interpretation of the Nash equilibrium already in his original formulation. Under his so called "mass-action" interpretation:

> "It is unnecessary to assume that the participants have full knowledge of the total structure of the game, or the ability and inclination to go through any complex reasoning processes. But the participants are supposed to accumulate empirical information on the relative advantages of the various pure strategies at their disposal."

Evolutionary explanations actually have an even longer tradition in economics, and can be found in the social sciences much before Darwin. However, the formal machinery for studying evolution did not start to develop until the 1970-ies, when two British biologists attacked the problem. Maynard Smith and Price (1973), and Maynard Smith (1974) model a biological setting where individuals from a large population are repeatedly drawn at random to play a symmetric two-player game, where payoff measures fitness. These individuals do not choose a strategy on rationalistic grounds, but are instead genetically programmed to play a specific strategy. The authors define a strategy as *evolutionarily stable* (ESS) if a homogeneous population of individuals, who all use this strategy, cannot be invaded by mutants using a different strategy. More specifically, an ESS is a strategy, such that if used by a sufficiently large population share, then it gives a higher expected payoff than any other strategy used by the rest of the population. An ESS is always in Nash equilibrium with itself, but a Nash equilibrium strategy is not necessarily an ESS. In this sense, the ESS concept is a refinement. For

[4] See e.g. Tan and Werlang (1988) and Aumann and Brandenburger (1995).

example, both pure strategies in Figure 0.1 are ESS:s, but the mixed Nash equilibrium strategy is not.

ESS is a static concept, but several authors, starting with Taylor and Jonker (1978), have also constructed dynamic models of evolutionary selection, such as the *replicator dynamics.* The idea here is that the population shares of individuals with different strategies increase or decrease at a rate depending on their current relative payoffs. The biological motivation is that a higher fitness should result in a higher rate of reproduction. By specifying the relationship between payoffs and growth rates of strategies, a dynamic system that can be solved for possible stable points or sets is obtained, and the points or sets to which the system will converge in the long run can be studied. For many such selection dynamics, there is a strikingly close connection between stable strategies or sets of strategies, and static concepts such as that of ESS and Nash equilibrium.

Applying evolutionary game theory directly to economics is not without problems. In economics, payoffs do generally not represent fitness, but profits or utility. Moreover, the selection mechanism is not biological, but the result of conscious choices by the individuals. To deal with this, economists have generalized the evolutionary selection dynamics to more or less rational adjustment processes. There is also an entire strand of evolutionary game theory, called *learning in games,* devoted to explicit models of how individuals adapt and change their strategies based on past experience.[5]

Learning has been extensively studied within the field of *decision theory*, where a single decision maker facing an exogenous and uncertain environment needs to make a decision based on some empirical distribution of outcomes. A notable example of learning in this context is *rational* or *Bayesian* learning. When learning is applied to game theory, matters become more complex, since each player's action may affect his opponents' future actions and beliefs – in other words, the environment is no longer exogenous. In game theory, a *learning rule* is a decision rule which prescribes a particular strategy for each possible history of all players' previous actions. The objective of the analysis of learning in games is generally to predict which strategy, or strategies, will be played in the limit, as time goes to infinity, by individuals using a particular learning rule. Models of *social learning* focus on the aggregate behavior of a population, whereas models of *individual learning* concentrate on the behavior of a particular individual. There is a close connection between models of learning in games and dynamic models of evolutionary selection. Several authors have shown that the dynamics of many learning models can be closely approximated by evolutionary selection dynamics.

[5]For an introduction, see Fudenberg and Levine (1998).

Three of the most well-known classes of learning rules in games, all with some experimental support (see Camerer and Ho, 1999), are *belief learning*, *reinforcement learning*, and *imitation*. Belief learners use the empirical distribution of past play to predict the distribution of their opponents' actions, and then play an optimal reply to this distribution. Reinforcement learners, on the other hand, only consider their own past play, and choose the strategy that has been most successful in the past. Imitators are similar to reinforcement learners, but instead of observing their own past play, they consider the past play of other individuals, and adopt the most successful strategy.

Apart from the unrealistic assumptions noted above, there is a second problem with the rationalistic paradigm of traditional game theory: the multiplicity of equilibria in many games. A multitude of different, more or less demanding, rationalistic equilibrium concepts have been proposed for dealing with this problem, but there is no consensus on which concept to use. Although several of these work well for some classes of games, they either have no cutting power, or give counterintuitive results in others. A major motivation for static and dynamic evolutionary concepts, and the theory of learning in games, is their ability to select among multiple equilibria in an intuitive manner. The set of evolutionarily stable strategies and the set of strategies to which an evolutionary selection dynamic or a learning rule converges, are often smaller than the full set of Nash equilibria of the game, as in the game in Figure 0.1 above.

In spite of this, there are many games where more than one equilibrium is selected by evolutionary and learning methods. Initial conditions, such as the initial shares of different strategies in the population or the initial history of actions, often determine the point to which the system will converge. In order to further sharpen the selective ability in such games, Foster and Young (1990) have developed an ingenious technique, which I use extensively in this thesis. This technique can be applied to games where evolution is modeled as a stochastic process. By perturbing this process, for instance by assuming that individuals with a small probability experiment or make mistakes and play a strategy at random instead of the one prescribed by their learning rule, the process can be made ergodic – i.e. independent of initial conditions. This implies that in the very long run, and for small experimentation probabilities, the process will spend almost all the time in a particular equilibrium which, in this sense, is selected. In the Coordination Game in Figure 0.1, this technique generally selects one of the two pure equilibria.

0.2 Summary of the Essays

The first two essays of this thesis analyze which behaviors will evolve when populations of individuals, using identical learning rules, play a finite game for a long time. In the first essay, the learning rule consists of imitating the most successful strategy in a sample of past play. This learning rule belongs to the above described class of imitation. In the second essay, the learning rule prescribes play of a *better reply* to a sample distribution of past play. This rule belongs to the above described class of belief learning. The third essay extends this setting by allowing for heterogeneous populations of individuals with both of these rules present.

The final essay goes one step further and considers evolutionary selection among a wide class of learning rules containing belief learning and reinforcement learning, as well as intermediate learning rules. More specifically, it asks whether there exists a learning rule, such that if used by a homogeneous population, then the population cannot be invaded by a mutant learning rule.

The framework in the first three essays is that of Young (1993, 1998). In each period, one individual from each of n populations, one for each player role, is drawn at random to play a finite n-player game. Each individual chooses a pure strategy according to a specific learning rule after observing a sample drawn without replacement from a finite history of play. With some probability, the individuals also make mistakes or experiment and play a strategy at random. Together, these assumptions define an ergodic Markov chain on the space of histories.

In the final essay, the setting is slightly different. In each of a finite number of periods, all members of a large population with an even number of individuals are randomly matched in pairs to play a two-player game. The individuals all use a learning rule in the class of rules that can be represented by Camerer and Ho's (1999) experimental model.

0.2.1 Stochastic Imitation in Finite Games

In the first essay, which is coauthored with Alexander Matros, it is assumed that all individuals are *imitators* who sample population-specific strategy and payoff realizations, and imitate the most successful behavior. In finite n-player games, we prove that, for a sufficiently low ratio between the sample and history size, only pure-strategy profiles in certain *minimal closed sets under the better-reply graph* will be played with positive probability in the limit, as the probability of mistakes tends to zero. If, in addition, the sample size is sufficiently large and the strategy profiles in one such minimal set have strictly higher payoffs than all other strategy profiles, then only the strategies in this set will be played with positive probability in the limit, provided that the minimal set is a product set. Applied to 2×2 Coordination Games, the Pareto dominant equilibrium is selected for a sufficiently large sample size.

0.2.2 Stochastic Better-Reply Dynamics in Finite Games

In the second essay, all individuals use a learning rule which prescribes play of a *better reply* to a sample distribution of past play. The *better-reply correspondence* maps distributions over the player's own and her opponents' strategies to those pure strategies giving the player a weakly better expected payoff. In finite n-player games, I prove that, for a sufficiently low ratio between the sample and history size only pure-strategy profiles in certain *minimal closed sets under better replies (MCUBR sets)* will be played with positive probability in the limit, as the probability of mistakes tends to zero. This result is consistent with and extends previous results on the equivalence of asymptotically stable sets and closed sets under better replies for a large class of deterministic selection dynamics in continuous time. Applied to 2×2 Coordination Games, the risk-dominant equilibrium is selected for a sufficiently large sample.

0.2.3 Stochastic Adaptation in Finite Games Played by Heterogeneous Populations

The third essay models stochastic learning in finite n-player games played by *heterogeneous* populations of imitators, better repliers, and also best repliers, who plays a best reply to a sample distribution of past play. In finite n-player games, I prove that, for a sufficiently low ratio between the sample and history size, only pure-strategy profiles in certain MCUBR sets will be played with positive probability in the limit, as the probability of mistakes tends to zero. If, in addition, the strategy profiles in one such MCUBR set have strictly higher payoffs than all other strategy profiles and the sample size is sufficiently large, then the strategies in this set will be played with probability one in the limit, as the probability of mistakes tends to zero. Applied to 2×2 Coordination Games, the Pareto dominant equilibrium is selected for sufficiently large samples, but in all symmetric and many asymmetric games, the risk dominant equilibrium is selected for sufficiently small sample size.

0.2.4 A Numerical Analysis of the Evolutionary Stability of Learning Rules

In the final essay, I define an evolutionary stability criterion for learning rules. Using Monte Carlo simulations, this criterion is then applied to a class of learning rules that can be represented by Camerer and Ho's (1999) model of learning. This class contains perturbed versions of two of the most well-known learning rules, reinforcement and belief learning, as special cases. The simulations show that belief learning is the only learning rule which is evolutionarily stable in almost all cases, whereas reinforcement learning is unstable in almost all cases. It is also found that in certain games, the stability of intermediate learning rules hinges critically on a parameter of the model and on relative payoffs.

Bibliography

[1] Alchian, A. (1950): "Uncertainty, Evolution and Game Theory," *Journal of Political Economy* 58, 211-222.

[2] Aumann, R., and A. Brandenburger (1995): "Epistemic Conditions for Nash Equilibria," *Econometrica* 63, 1161-1180.

[3] Friedman, M. (1953): "The Methodology of Positive Economics," in M. Friedman, *Essays in Positive Economics.* University of Chicago Press.

[4] Foster, D., and P. Young (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* 38, 219-32.

[5] Fudenberg, D., and D. Levine (1998): *The Theory of Learning in Games.* The MIT Press.

[6] Maynard Smith, J., and G. R. Price (1973): "The Logic of Animal Conflict," *Nature* 246, 15-18.

[7] Maynard Smith, J. (1974): "The Theory of Games and the Evolution of Animal Conflicts," *Journal of Theoretical Biology* 47, 209-221.

[8] Myerson R. B. (1991): *Game Theory. Analysis of Conflict.* Harvard University Press.

[9] Nash, J. (1950). "Non Cooperative Games," Ph.D. thesis, Princeton University.

[10] Ritzberger, K. (2001): *Foundations of Non-Cooperative Game Theory.* Oxford University Press.

[11] Samuelson, L. (1997): *Evolutionary Games and Equilibrium Selection.* The MIT Press.

[12] Tan, T., and S. R. Werlang (1988): "The Bayesian Foundations of Solution Concepts in Games," *Journal of Economic Theory* 45, 370-391.

[13] Taylor, P., and L. Jonker (1978): “Evolutionary Stable Strategies and Game Dynamics,” *Mathematical Biosciences* 40: 145-56.

[14] Vega-Redondo, F. (1996): *Evolution, Games, and Economic Behavior.* Oxford University Press.

[15] —— (2001): *Economics and the Theory of Games.* Cambridge University Press, forthcoming.

[16] Weibull, J. W. (1997): "What Have We Learned from Evolutionary Game Theory So Far?," Working paper no. 487, The Research Institute of Industrial Economics.

[17] —— (1995): *Evolutionary Game Theory.* The MIT Press.

[18] Young, P. H. (1993). “The Evolution of Conventions,” *Econometrica* 61, 57-84.

[19] —— (1998). *Individual Strategy and Social Structure.* Princeton University Press.

Chapter 1

Stochastic Imitation in Finite Games[1]

1.1 Introduction

In most game-theoretical models of learning, the individuals are assumed to know a great deal about the structure of the game, such as their own payoff function and all players' available strategies. However, for many applications, this assumption is neither reasonable nor necessary; in many cases, individuals may not even be aware that they are playing a game. Moreover, equilibrium play may be achieved even with individuals who have very little knowledge of the game, an observation made already in 1950 by John F. Nash. In his unpublished Ph.D. thesis (1950), he referred to it as "the 'mass-action' interpretation of equilibrium points." Under this interpretation:

> "It is unnecessary that the participants have full knowledge of the total structure of the game, or the ability and inclination to go through any complex reasoning processes. But the participants are supposed to accumulate empirical information on the relative advantages of the various pure strategies at their disposal."

In the present paper, we develop a model in this spirit, where individuals are only required to know their own available pure strategies and a sample of the payoffs that a subset of these strategies have earned in the past. In spite of this weak assumption, our model predicts equilibrium play in many games. We use an evolutionary framework with perpetual random shocks

[1] Coauthored with Alexander Matros.

similar to Young (1993), but our assumption of individual behavior is different. Whereas the individuals in his model play a myopic best reply to a sample distribution of their opponents' strategies, the individuals in our model imitate other individuals in their own population. Imitation is a behavior with both experimental, empirical, and theoretical support.[2]

More specifically, we assume that in every period, individuals are drawn at random from each of n arbitrary-size populations to play a finite n-player game. Each of these individuals observes a sample from a finite history of her population's past strategy and payoff realizations. Thereafter, she imitates by choosing the most attractive strategy in her sample. This could, for instance, be the strategy with the highest average payoff, or that with the highest maximum payoff.[3,4] With some small probability, the individuals also make errors or experiment, and instead choose any strategy at random from their set of strategies.[5] Altogether, this results in an ergodic Markov process, which we denote *imitation play*, on the space of histories. We study the stationary distribution of this process as the experimentation probability tends to zero.

Imitation in a stochastic setting has previously been studied by Robson and Vega-Redondo (1996), who modify the framework of Kandori *et al.* (1993) to allow for random matching. More precisely, they assume that in each period, individuals are randomly matched for a finite number of rounds and tend to adopt the strategy with the highest average payoff across the population. Robson and Vega-Redondo (1996) assume either single- or two-population structures and obtain results for symmetric 2×2 games and two-player games of common interest.

However, our model differs considerably from this and other stochastic learning models, and has several advantages. First, we are able to prove general results, applicable to any finite n-player game, about the limiting distribution of imitation play. We are thus not restricted to the two classes

[2] For experimental support of imitation, see for example, Huck *et al.* (1999, 2000) and Duffy and Feltovich (1999), for empirical support see Graham (1999), Wermers (1999), and Griffiths *et al.* (1998), and for theoretical support, see Björnerstedt and Weibull (1996) and Schlag (1998, 1999).

[3] In the special case when each population consists of only one individual, this behavior can be interpreted as a special kind of reinforcement learning.

[4] This behavior is related to one of the interpretations of individual behavior in Osborne and Rubinstein (1998), where each individual first samples each of her available strategies once and then chooses the strategy with the highest payoff realization.

[5] An alternative interpretation, which provides a plausible rationale for experimentation and is consistent with the knowledge of individuals in the model, is that if and only if the sample does not contain all available strategies, then with a small probability, the individual instead picks a strategy not included in the sample at random.

of games in Robson and Vega-Redondo (1996), or even to a generic class of games, as in Young (1998). Second, we find that this distribution has some interesting properties. For instance, it puts probability one on an efficient set of outcomes in a large class of n-player games. Third, the speed of convergence of our process is relatively high. We show that in 2×2 Coordination Games, the expected first passage time may be considerably shorter than in Young (1993), Kandori *et al.* (1993), and Robson and Vega-Redondo (1996), for small experimentation probabilities.

The perturbed version of imitation play is a regular perturbed Markov process. This implies that the methods employed by Young (1993) can be used to calculate the states that will be played with positive probability by the stationary distribution of the process as the experimentation probability tends to zero, i.e. the *stochastically stable states.* We prove three results which facilitate this calculation and enable us to characterize the set of such states. These results hold in finite n-player games, provided the information available to the individuals is sufficiently incomplete and their sample size is sufficiently large.

First, we show that from any initial state, the unperturbed version of imitation play converges to a state which is a repetition of a single pure-strategy profile, a *monomorphic state.* Hence, the stochastically stable states of the process belong to the set of monomorphic states.

Second, we prove that in the limit, as the experimentation probability tends to zero in the perturbed process, only pure-strategy profiles in particular subsets of the strategy-space are played with positive probability. These sets, which we denote *minimal constructions,* are defined in the following way. By drawing directed edges from each pure-strategy profile to the pure-strategy profiles which give exactly one of the players at least as high payoff, a *better-reply graph* is obtained. We define a *construction* as a set of pure-strategy profiles from which there are no such outgoing edges. A minimal construction is a minimal set with this property. Minimal constructions are similar to Sobel's (1993) definition of *non-equilibrium evolutionary stable* (*NES*) sets for two-player games and to what Nöldeke and Samuelson (1993) call *locally stable components* in their analysis of extensive form games. They are also closely related to *minimal closed sets under better replies* (Ritzberger and Weibull, 1995). We show that every minimal set closed under better replies contains a minimal construction and that if a minimal construction is a product set, then it is also a minimal set closed under better replies. The relationship between minimal constructions and the limiting distribution of imitation play should be contrasted with Young's (1998) finding that adaptive play for generic games selects pure-strategy profiles in minimal closed sets under best replies.

Finally, we show that in a certain class of games, imitation play selects efficient outcomes. If the pure-strategy profiles in a minimal construction have strictly higher payoffs than all other pure-strategy profiles and the sample size is sufficiently large, then the pure-strategy profiles in this set will be played with probability one in the limit, provided that the minimal construction is a product set. This is a generalization of Robson and Vega-Redondo's (1996) result that under certain conditions a Pareto-dominant pure-strategy profile corresponds to a unique stochastically stable state in two-player games of common interest.

Applied to 2×2 games, these three results give clear predictions. In Coordination Games, where one equilibrium strictly Pareto dominates the other, imitation play selects the strictly Pareto-superior Nash-equilibrium. This result differs sharply from the predictions in Young's (1993) and Kandori *et al.'s* (1993) models, where the stochastically stable states correspond to the risk-dominant equilibria, but it is consistent with the predictions of Robson and Vega-Redondo's (1996) model for symmetric Coordination Games. However, if neither equilibria Pareto dominates the other, the latter model may select the risk-dominant equilibrium, whereas both equilibria are played with positive probability in our model. In games without pure Nash equilibria, all four monomorphic states are stochastically stable.

The paper is organized as follows. In Section 2, we define the unperturbed and perturbed versions of imitation play. In Section 3, we derive general results for the limiting distribution of the process. In Section 4, we investigate some of the properties of our solution concept, the minimal construction. In Section 5, we apply our results to 2×2 games and compare our findings to those in previous literature. In Section 6, we discuss an extension of the model and in Section 7, we conclude. Omitted proofs can be found in the Appendix.

1.2 Model

The model described below is similar to Young (1993), but the sampling procedure is modified and individuals employ a different decision-rule. Let Γ be a n-player game in strategic form and let X_i be the finite set of pure strategies x_i available to player $i \in \{1, ..., n\} = N$. Define the product sets $X = \prod_i X_i$ and $\square(X) = \prod_i \Delta(X_i)$, where $\Delta(X_i) = \{p \in \mathbb{R}_+^{|X_i|} \mid p'\mathbf{1} = 1\}$ is the simplex of individual strategies. Let $C_1, ..., C_n$ be n finite and non-empty populations of individuals. These populations need not be of the same size, nor need they be large. Each member of population C_i is a candidate to play role i in the game Γ. All individuals in population C_i have payoffs represented by the utility function $\pi_i : X \to \Pi_i$, where $\Pi_i \subset \mathbb{R}$.[6] Expected payoffs are represented by the function $u_i : \square(X) \to \mathbb{R}$. Note that we write "players" when referring to the game Γ and "individuals" when referring to the members of the populations.

The play proceeds as follows. Let $t = 1, 2, ...$ denote successive time periods. The game Γ is played once every period. In period t, one individual is drawn at random from each of the n populations and assigned to play the corresponding role. An individual in role i chooses a pure strategy x_i^t from a subset of her strategy space X_i, according to a rule that will be defined below. The pure-strategy profile $x^t = (x_1^t, ..., x_n^t)$ is recorded and referred to as *play* at time t. The *history* of plays up to time t is the sequence $h^t = (x^{t-m+1}, ..., x^t)$, where m is a given positive integer, the *memory size* of all individuals.

Let h be an arbitrary history. Denote a sample of s $(1 < s \leq m)$ elements from the m most recent strategy choices by individuals in population C_i by $w_i \in X_i^s$, and the corresponding payoff realizations by $v_i \in \Pi_i^s$. For any history h, the *maximum average correspondence*, $\alpha_i : X_i^s \times \Pi_i^s \twoheadrightarrow X_i$, maps each pair of strategy sample w_i and payoff sample v_i to the pure strategy (or the set of strategies) with the highest average payoff in the sample. Following Young (1993), we can think of the sampling process as beginning in period $t = m + 1$ from some arbitrary initial sequence of m plays h^m. In this period and every period thereafter, each individual in player role i inspects a pair (w_i, v_i) and plays a pure strategy $x_i \in \alpha_i(w_i, v_i)$. This defines a finite Markov process on the finite state space $H = X^m$ of histories. Given a history $h^t = (x^{t-m+1}, ..., x^t)$ at time t, the process moves to a state of the form $h^{t+1} = (x^{t-m+2}, ..., x^t, x^{t+1})$ in the next period. Such a state is called a *successor* of h^t. The process moves from the current state h to a successor state h' in each period according to the following transition rule. For each $x_i \in X_i$, let $p_i(x_i \mid h)$ be the probability that individual i chooses pure strategy x_i. We assume $p_i(x_i \mid h) > 0$ if and only if there exists a sample of population-specific strategy choices w_i and payoff outcomes v_i, such that $x_i \in \alpha_i(w_i, v_i)$. If x is the rightmost element of h', the probability of moving from h to h' is $P_{hh'}^{\alpha,m,s,0} = \prod_{i=1}^n p_i(x_i \mid h)$ if h' is a successor of h, and $P_{hh'}^{\alpha,m,s,0} = 0$, if h' is not a successor of h. We call the process $P^{\alpha,m,s,0}$ *imitation play with memory m and sample size s.*

As an example, consider imitation play with memory $m = 6$ and sample size $s = 3$ in the 2×3 game in Figure 1.1.

[6] Actually, utility functions need not be identical within each population for any of the results in this paper. It is sufficient if each individual's utility function is a positive affine transformation of a population-specific utility function.

	a	b	c
A	2, 2	0, 0	3, 3
B	0, 0	1, 1	0, 0

FIGURE 1.1

Let $h = ((A,a),(B,a),(A,b),(B,b),(A,c),(B,c))$ be the initial history. Assume that the individual in the role of the row player (player 1) draws the last three plays from this history, such that $w_1 = (B,A,B)$ and $v_1 = (1,3,0)$. This gives an average payoff of 3 to strategy A and 1/2 to strategy B. Hence, the individual in role of the row player will choose strategy A in the next period. Further, assume that the individual in the role of the column role (player 2) draws the first three plays, such that $w_2 = (a,a,b)$ and $v_2 = (2,0,0)$. This gives an average payoff of 1 to strategy a and 0 to strategy b. Strategy c cannot be chosen since it is not included in the sample. Hence, the individual in the column role will choose strategy a in the next period. Altogether, this implies that the unperturbed process will move to state $h' = ((B,a),(A,b),(B,b),(A,c),(B,c),(A,a))$ in the next period.

The perturbed process can be described as follows. In each period, there is some small probability $\varepsilon > 0$ that each individual i drawn to play chooses a pure strategy at random from X_i, instead of according to the imitation rule. The event that i experiments is assumed to be independent of the event that j experiments for every $j \neq i$ and across time periods. For every i, let $q_i(x_i \mid h)$ be the conditional probability that i chooses $x_i \in X_i$, given that i experiments and the process is in state h. We assume that $q_i(x_i \mid h) > 0$ for all $x_i \in X_i$. Suppose that the process is in state h at time t. Let $D \subset N$ be any subset of d player roles. The probability is $\varepsilon^d(1-\varepsilon)^{n-d}$ that exactly the individuals drawn to play the roles in D experiment and the others do not. Conditional on this event, the transition probability of moving from h to h' is $Q^D_{hh'} = \prod_{i \in D} q_i(x_i \mid h) \prod_{i \notin D} p_i(x_i \mid h)$, if h' is a successor of h and x is the rightmost element of h', and $Q^D_{hh'} = 0$, if h' is not a successor of h. This gives the following transition probability of the perturbed Markov process, similar to equation (2) in Young (1993, p. 67):

$$P^{\alpha,m,s,\varepsilon}_{hh'} = (1-\varepsilon)^n P^{\alpha,m,s,0}_{hh'} + \sum_{D \subset N, D \neq \emptyset} \varepsilon^{|D|}(1-\varepsilon)^{|N \setminus D|} Q^D_{hh'}. \qquad (1.1)$$

The process $P^{\alpha,m,s,\varepsilon}$ is denoted *imitation play with memory m, sample size s, and experimentation probabilities ε.*

1.3 Stochastic Stability in Finite Games

In this section, we turn our attention to the limiting distribution of imitation play as the experimentation probability tends to zero. We first show that we can apply some of the tools in Young (1993) to calculate this distribution. Thereafter, we prove that it puts positive probability only on the pure-strategy profiles in particular subsets of the strategy-space.

1.3.1 Preliminaries

In what follows, we will make use of the following definitions. A *recurrent class* of the process $P^{\alpha,m,s,0}$ is a set of states such that there is zero probability of moving from any state in the class to any state outside, and there is a positive probability of moving from any state in the class to any other state in the class. We call a state h' *absorbing* if it constitutes a singleton recurrent class and we refer to a state $h_x = (x, x, ..., x)$, where x is any pure-strategy profile from X, as a *monomorphic state.* In other words, a monomorphic state is a state where the individuals in each player role played the same pure strategy in the last m periods. If each player i has $|X_i| \geq 1$ strategies in the game Γ, then there are $|X| = \prod_i |X_i|$ monomorphic states in this game. The following result shows that monomorphic states correspond one to one with the recurrent classes of imitation play.

Theorem 1 *All monomorphic states are absorbing states of the unperturbed process $P^{\alpha,m,s,0}$. If $s/m \leq 1/2$, the process converges to a monomorphic state with probability one.*

PROOF: It is evident that for s such that $1 \leq s \leq m$, any monomorphic state is an absorbing state, since any sample from a monomorphic state will contain one strategy only. We shall prove that if $s/m \leq 1/2$, then the monomorphic states are the only recurrent classes of the unperturbed process. Consider an arbitrary initial state $h^t = (x^{t-m+1}, ..., x^t)$. If $s/m \leq 1/2$, there is a positive probability that all individuals drawn to play sample from $x^{t-s+1}, ..., x^t$ (i.e. i, for $i \in N$, sample $x_i^{t-s+1}, ..., x_i^t$ and $\pi_i^{t-s+1}, ..., \pi_i^t$) in every period from $t+1$ to $t+s$ inclusive. All of them play the pure strategy with the highest average payoff in their sample. Without loss of generality, assume that this is a unique pure strategy x_i^* for each of the player roles (if there is more than one pure strategy, all of them have positive probability according to the assumptions). With positive probability, all the individuals drawn to play thereafter sample only from plays more recent than x^t in every period from $t+s+1$ to $t+m$ inclusive. Since all of these samples have the form $w_i^* = (x_i^*, ..., x_i^*)$ and $v_i^* = (\pi_i^*, ..., \pi_i^*)$, the unique pure strategy with the highest payoff in the sample is x_i^*. Hence, there is a positive probability of at time $t+m$ obtaining a history $h^{t+m} = (x^*, ..., x^*)$, a monomorphic state. It follows that for $s/m \leq 1/2$, the only recurrent classes of the unperturbed process are the monomorphic states. *Q.E.D.*

Lemma 1 below implies that Theorem 3.1 in Young (1998, p. 56) applies to imitation play. With a slightly modified notation, this theorem is referred to as Theorem 2 in this paper.

Definition 1 *(Young, 1998) P^ε is a* ***regular perturbed Markov process*** *if P^ε is irreducible for every $\varepsilon \in (0, \varepsilon^*]$, and for every states h, h', $P^\varepsilon_{hh'}$ approaches $P^0_{hh'}$ at an exponential rate, i.e. $\lim_{\varepsilon \to 0} P^\varepsilon_{hh'} = P^0_{hh'}$ and if $P^\varepsilon_{hh'} > 0$ for some $\varepsilon > 0$, then $0 < \lim_{\varepsilon \to 0} \frac{P^\varepsilon_{hh'}}{\varepsilon^{r_{hh'}}} < \infty$ for some $r_{hh'} \geq 0$. The real number $r_{hh'}$ is called the* ***resistance*** *of the transition from h to h'.*

Lemma 1 *Imitation play is a regular perturbed Markov process.*

PROOF: See the Appendix.

Definition 2 *(Foster and Young, 1990) Let μ^ε be the unique stationary distribution of an irreducible process, P^ε. A state h is* ***stochastically stable*** *if* $\lim_{\varepsilon \to 0} \mu^\varepsilon(h) > 0$.

Let an unperturbed Markov process P^0 with $s/m \leq 1/2$ have recurrent classes/ monomorphic states $h_{x^1}, ..., h_{x^{|X|}}$ (according to Theorem 1, there is a one to one correspondence between the recurrent classes and the monomorphic states for $s/m \leq 1/2$). The following concepts are due to Freidlin and Wentzell (1984) and Foster and Young (1990). For each pair of distinct recurrent classes, an *xy-path* is a sequence of states $\zeta = (h_x, ..., h_y)$ beginning in h_x and ending in h_y. The *resistance* of this path is the sum of the resistances on the edges that compose it. Let r_{xy} be the least resistance over all xy-paths. Construct a complete directed graph with $|X|$ vertices, one for each recurrent class. The weight on the directed edge $h_x \to h_y$ is r_{xy}. A *tree* rooted at h_x is a set of $|X| - 1$ directed edges such that, from every vertex different from h_x, there is a unique directed path in the tree to h_x. The *resistance of such a rooted tree* $\Im(x)$ is the sum of the resistances $r_{x'x''}$ on the $|X| - 1$ edges that compose it. The *stochastic potential* $\rho(x)$ of a recurrent class h_x is the minimum resistance over all trees rooted at h_x.

Theorem 2 *(Young, 1998) Let P^ε be a regular perturbed Markov process and let μ^ε be the unique stationary distribution of P^ε for $\varepsilon > 0$. Then,* $\lim_{\varepsilon \to 0} \mu^\varepsilon =$

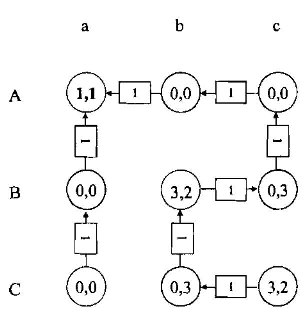
μ^0 exists, and is a stationary distribution of P^0. The stochastically stable states are precisely those states contained in the recurrent classes of P^0 having minimum stochastic potential.

In order to illustrate how to calculate the stochastic potential under imitation play, we present an example of a two-player game. In the game in Figure 1.2, every player has three strategies, labeled A, B and C for the first player and a, b and c for the second player. The game has one strict Nash equilibrium (A, a), where both players gain less than in a mixed equilibrium with the probability mixture $1/2$ on B (b) and $1/2$ on C (c) for the first (second) player.[7]

	a	b	c
A	$1,1$	$0,0$	$0,0$
B	$0,0$	$3,2$	$0,3$
C	$0,0$	$0,3$	$3,2$

FIGURE 1.2

The monomorphic states $h_{(B,b)}$ and $h_{(C,c)}$ are Pareto superior to monomorphic state $h_{(A,a)}$. Denote by $x_1 \in \{A, B, C\}$ some strategy choice by player 1 and $x_2 \in \{a, b, c\}$ some strategy choice by player 2. To find the stochastically stable monomorphic states, construct directed graphs with nine vertices, one for each monomorphic state. In Figure 1.3, we illustrate two such trees. The numbers in the squares correspond to the resistances of the directed edges and the numbers in the circles represent the payoffs associated with the monomorphic states. It is easy to check that for $s > 2$, $\rho(A, a) = 8$, whereas all other monomorphic states have a stochastic potential of 9. Hence, monomorphic state $h_{(A,a)}$ is *stochastically stable.*



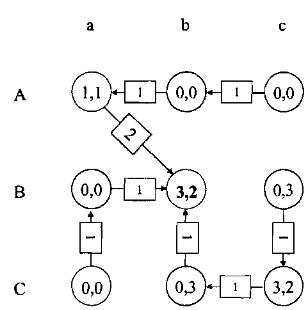


FIGURE 1.3–Minimum-resistance trees rooted at $h_{(A,a)}$ and $h_{(B,b)}$, respectively.

[7]There is also a third equilibrium, $((\frac{5}{7}, \frac{1}{7}, \frac{1}{7}), (\frac{3}{5}, \frac{1}{5}, \frac{1}{5}))$.

1.3.2 Sets of Stochastically Stable States

In this subsection, we will show that the stochastically stable states of imitation play correspond to pure-strategy profiles in particular subsets of the strategy-space. In order to do this, we need to introduce some new concepts. Given an n-player game Γ with finite strategy space $X = \prod X_i$, associate each pure-strategy profile $x \in X$ with the vertex of a graph. Draw a directed edge from vertex x to vertex x', if and only if both of the following conditions hold:

i) there exists exactly one player i such that $x_i \neq x_i'$ and

$$u_i\left(x_i', x_{-i}\right) \geq u_i\left(x\right), \tag{1.2}$$

ii) there does not exist $x_i'' \neq x_i$ and $x_i'' \neq x_i'$, such that

$$u_i\left(x_i', x_{-i}\right) > u_i\left(x_i'', x_{-i}\right) > u_i\left(x\right). \tag{1.3}$$

Call the graph obtained in this manner the *better-reply graph* of game Γ.[8] A *better-reply path* is a sequence of the form $x^1, x^2, ..., x^l$, such that each pair (x^j, x^{j+1}) corresponds to a directed edge in the better-reply graph. A *sink* is a vertex with no outgoing edges. Clearly, x is a sink if and only if it is a strict Nash equilibrium. A better-reply graph can also contain directed cycles.

Figure 1.4 shows the better-reply graph for a game with two sinks - (C, a) and (B, b). The *basin* of a sink is the set of all vertices from which there exists a directed path ending at that sink. Note that a vertex may be in several basins simultaneously; for example, (B, a) is in the basin of (C, a) and also in the basin of (B, b).

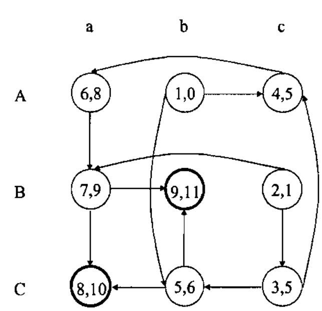


FIGURE 1.4–Better-reply graph.

[8]Condition ii) is included only for ease of exposition. The omission of this condition would complicate the notation considerably, without changing our results.

Lemma 2 *If x and x' are vertices of the better-reply graph of the finite n-player game Γ, and there exists a directed edge from vertex x to vertex x', then the following inequality holds for the stochastic potentials $\rho(x)$ and $\rho(x')$ of the monomorphic states h_x and $h_{x'}$, respectively:*

$$\rho(x) \geq \rho(x'). \tag{1.4}$$

PROOF: Suppose that the claim is false, such that $\rho(x) < \rho(x')$. Note that $\rho(x)$ is the minimum resistance over all trees rooted at state h_x. Construct a new tree rooted at $h_{x'}$ by taking (one of) the tree(s) with minimum resistance rooted at h_x, adding the directed edge from vertex x to vertex x' and deleting the directed edge from x'. The resistance of the added edge is exactly one and that of the deleted edge at least one, so the total resistance of the new tree is $\rho(x)$ at most, implying a contradiction. *Q.E.D.*

Definition 3 *A non-empty set of vertices V is said to be closed under the better-reply graph, or V is a* ***construction,*** *if there are no directed edges from any of the vertices in V to any vertex outside V. Such a set is called a* ***minimal construction*** *if it does not properly contain another construction.*

From the definition follows that every game contains a minimal construction. Moreover, any sink is a minimal construction. The game in Figure 1.5 has two minimal constructions: the sink (A, a), and the set $V = \{(B,b), (B,c), (C,c), (C,b)\}$, which can be considered as a single directed cycle $(B,b) \to (B,c) \to (C,c) \to (C,b) \to (B,b)$.

	a	b	c
A	1, 1	0, 0	0, 0
B	0, 0	3, 2	2, 3
C	0, 0	2, 3	3, 2

FIGURE 1.5

Note that our definition of minimal constructions in normal form games is related to Sobel's (1993) definition of *non-equilibrium evolutionary stable (NES)* sets in two-player games. Our concept is also similar to Nöldeke and Samuelson's (1993) definition of *locally stable components* in extensive form games.

Lemma 3 *Let V be a minimal construction of a finite n-player game Γ. Then, for any two vertices $x, x' \in V$, there exist better-reply paths $x, ..., x'$ and $x', ..., x$, which connect these vertices.*

PROOF: Suppose that the claim is false and there exist two vertices $x, x' \in V$ such that there is no better-reply path from x to x'. Consider all better-reply paths, starting at vertex x. Since X is finite, there exists a finite number of vertices along all these paths. Collect all these vertices. By construction, this set of vertices has only ingoing edges and by assumption, it does not contain the vertex x'. Hence, the constructed set of vertices is a minimal construction and a proper subset of the minimal construction V. This is a contradiction, so the claim must be true. *Q.E.D.*

To every vertex of a minimal construction corresponds a monomorphic state, which is a repetition of the associated pure-strategy profile, $x \in X$. Call the set of these monomorphic states a *minimal α-configuration.* We are now in a position to state the following main theorem.

Theorem 3 *Let Γ be a finite n-player game. If $s/m \leq 1/2$ and ε is sufficiently small, the perturbed process $P^{\alpha,m,s,\varepsilon}$ puts arbitrarily high probability on the minimal α-configuration(s) with minimum stochastic potential.*

PROOF: See the Appendix.

In order to prove this claim, we use Lemma 2 and Lemma 3 to show that: A) all monomorphic states in a minimal construction have equal stochastic potential and B) for every monomorphic state which does not belong to any minimal construction, there exists a monomorphic state with lower stochastic potential. The theorem establishes a relation between the stochastically stable states of imitation play and minimal constructions, which is similar to the relationship proved between the stochastically stable states of *adaptive play* and MCURB sets in Theorem 7.2 of Young (1998, p. 111).

We say that a finite set Y of pure-strategy profiles *strictly Pareto dominates* a pure-strategy profile x if for any pure-strategy profile $y \in Y$, $\pi_i(y) > \pi_i(x)$, for all i. The following theorem shows that imitation play selects sets of efficient outcomes in a large class of games.

Theorem 4 *Let Γ be a finite n-player game and suppose that there exists a minimal construction V, which is a product set and strictly Pareto dominates all pure-strategy profiles outside V. If $s/m \leq 1/2$, s is sufficiently large and ε is sufficiently small, the perturbed process $P^{\alpha,m,s,\varepsilon}$ puts arbitrarily high probability on the set of monomorphic states with span V.*

PROOF: See the Appendix.

The intuition behind this result is that for a sufficiently large sample size, the transition from a state inside V to any state outside V requires more mistakes than the number of player roles, while the opposite transition requires one mistake per player role at most. The following corollary follows immediately from Theorem 4.

Corollary 1 *Let Γ be a finite n-player game with a strict Nash Equilibrium x, that strictly Pareto dominates all other pure-strategy profiles. If $s/m \leq 1/2$ and s is sufficiently large, the monomorphic state h_x is a unique stochastically stable state.*

The requirement in Theorem 4 that V be a product set is necessary, as shown by the game in Figure 1.6.

	a	b	c
A	$1,1$	$0,0$	$0,0$
B	$0,0$	$0,0$	$5,3$
C	$0,0$	$2,4$	$5,4$

FIGURE 1.6

In this game, the minimal construction $V = \{(C,b), (C,c), (B,c)\}$ strictly Pareto dominates all pure-strategy profiles outside V. It is evident that two mistakes are enough to move from monomorphic state $h_{(A,a)}$ to any monomorphic state in the minimal construction V. We will show that two mistakes are also enough to move from the minimal construction V to an outside state. Suppose the process is in the state $h_{(C,c)}$ at time t. Further, suppose that the individual in player role 1 plays B instead of C at time $t+1$ by mistake. This results in play (B,c) at time $t+1$. Assume that the individual in player role 2 makes a mistake and plays b instead of c, and that the individual in player role 1 plays C in period $t+2$. Hence, the play at time $t+2$ is (C,b). Assume that the individuals in both player roles sample from period $t-s+2$ to period $t+2$ for the next s periods. This means that the individuals in player role 1 choose to play B and the individuals in player role 2 choose to play b from period $t+3$ to period $t+s+2$. There is a positive probability that from period $t+s+3$ through period $t+m+2$, the individuals in both player roles will sample from periods later than $t+2$. Hence, by the end of period $t+m+2$, there is a positive probability that the process will have reached the monomorphic state $h_{(B,b)}$ outside the minimal construction V. It is now straightforward to show that all the monomorphic states $h_{(A,a)}$, $h_{(C,b)}$, $h_{(C,c)}$, and $h_{(B,c)}$ have equal stochastic potential.

1.4 Properties of Minimal Constructions

In this section, we investigate the relationship between minimal constructions and other set-wise concepts. We also analyze whether strictly dominated strategies can be included in minimal constructions and possible invariance under payoff transformations.

1.4.1 Relation to Minimal Closed Sets Under Better Replies

We start by the relation between minimal constructions and so-called minimal closed sets under better replies. It might not be surprising that there is quite a strong connection between these two concepts.

Definition 4 *(Ritzberger and Weibull, 1995) Let the* ***better-reply correspondence*** $\gamma = \times_{i \in N} \gamma_i : \square(X) \twoheadrightarrow X$ *be defined as follows*

$$\gamma_i(p) = \{x_i \in X_i \mid u_i(x_i, p_{-i}) \geq u_i(p)\}, \forall i \in N. \tag{1.5}$$

A product set of strategies is a set of form $Y = \Pi_i Y_i$, where each Y_i is a non-empty subset of X_i. Let $\Delta(Y_i)$ denote the set of probability distributions over Y_i, and let $\square(Y) = \Pi_i \Delta(Y_i)$ denote the product set of such distributions. Let $\gamma_i(Y)$ denote the set of strategies in X_i that are better replies by i to some distribution $p \in \square(Y)$ and let $\gamma(Y) = \Pi_i \gamma_i(Y)$.

Definition 5 *A non-empty product set $Y \subset X$ is said to be* ***closed under better replies*** *(or Y is a* ***CUBR*** *set) if $\gamma(Y) \subset Y$. A set is called a* ***minimal closed set under better replies (MCUBR)*** *set if it is CUBR and contains no proper subset which is CUBR.*

Proposition 1 *Every MCUBR set contains a minimal construction. If a minimal construction is a product set, then it coincides with a MCUBR set.*

PROOF: We start from the first claim. Let $Y = \Pi_i Y_i$ be a MCUBR set. Associate each pure-strategy profile $x \in X$ with the vertex of a graph and draw a better-reply graph for game Γ on the finite strategy space X. By the definition of a MCUBR set, there are no directed edges from the set of vertices in Y to any vertices outside Y in the better-reply graph. This means that the set of vertices in Y is a minimal construction or a construction containing a minimal construction.

The last claim follows immediately from the definitions of a MCUBR set and a minimal construction. *Q.E.D.*

Combined with Theorem 3, this implies that if all minimal constructions of a game are product sets, then imitation play converges to the monomorphic states of MCUBR sets. Note that every minimal construction is not necessarily included in a MCUBR set. Consider the game in Figure 1.7.

	a	b	c
A	2,2	1,1	0,0
B	1,1	0,0	3,3
C	0,0	3,3	3,3

FIGURE 1.7

In this game, there is a unique MCUBR set, $\{(A, a)\} = Y$, and two minimal constructions, $V_1 = \{(A, a)\}$ and $V_2 = \{(B, c), (C, b), (C, c)\}$. Hence, V_2 is not contained in any MCUBR set.

1.4.2 Relation to CURB Sets

The minimal CURB set is another set-wise concept. Young (1998) shows that there is a strong relation between minimal CURB sets and stochastically stable states in his model. Let $Z = \Pi_{i=1}^{n} Z_i$ be a product set of strategies, and let $\beta_i(Z_{-i})$ denote the set of strategies in X_i that are best replies by i to some distribution $p_{-i} \in \square(Z_{-i})$. Define $\beta(Z) = \Pi_{i=1}^{n} \beta_i(Z_{-i})$.

Definition 6 *(Basu and Weibull, 1991) A non-empty product set $Z \subset X$ is said to be* ***closed under rational behavior*** *(or Z is a* ***CURB*** *set) if $\beta(Z) \subset Z$. Such a set is called a* ***minimal closed set under rational behavior (MCURB)****, if it does not properly contain a CURB set.*

It is straightforward to show that $\beta(Z) = Z$ for any MCURB set Z. Let the *span* of a non-empty set of vertices V, denoted by $S(V)$, be the product set of all strategies appearing in some pure-strategy profile of V. By the following two examples, we will show that a MCURB set and (the span of) a minimal construction are different set-wise concepts. In the game in Figure 1.8, the MCUBR set, the minimal construction and the MCURB set are all different.

	a	b	c
A	1,3	3,2	3,1
B	3,2	2,3	2,1
C	2,3	1,1	1,2

FIGURE 1.8

In this game, $\{A, B, C\} \times \{a, b, c\} = Y$ is a MCUBR set. There is a unique minimal construction: $V = \{(A, a), (C, a), (B, a), (B, b), (A, b)\}$ with span $S(V) = \{A, B, C\} \times \{a, b\}$. It is obvious that $\{A, B\} \times \{a, b\} = Z$ is the only MCURB set. Hence, $S(V) \cap Y = S(V)$ and $S(V) \cap Z = Z$ in this game.

One may ask whether the MCURB set is always included in the span of a minimal construction. In other words, is the set-wise relationship $S(V) \cap Z = Z$ true in general? The answer is no. Consider the game in Figure 1.9.

	a	b	c
A	3, 1	−10, 0	1, 3
B	−10, 0	1, 3	3, 1
C	1, 3	3, 1	−10, 0
D	0, 0	0, 0	0, 0

FIGURE 1.9

In this game, A is a best reply to a, a is a best reply to C, C is a best reply to b, b is a best reply to B, B is a best reply to c, and c is a best reply to A. Hence, any CURB set involving one or more of $\{A, a, C, b, B, c\}$ must involve them all. Hence, it must involve D, because D is a best reply to the probability mixture $1/3$ on a, $1/3$ on b,and $1/3$ on c. We conclude that $\{A, B, C, D\} \times \{a, b, c\} = Z$ is the unique MCURB set. However, the unique minimal construction is $V = \{(A, a), (C, a), (C, b), (B, b), (B, c), (A, c)\}$ with span $S(V) = \{A, B, C\} \times \{a, b, c\}$, such that $S(V) \cap Z = S(V) \neq Z$.

1.4.3 Relation to Strictly Dominated Strategies

Our model assumes that individuals are boundedly rational. In this subsection, we ask whether any dominated strategies will be played in the stochastically stable states of imitation play. Consider a 2×2 game where strategy A strictly dominates strategy B for the row player. In such a game, the minimal construction will either consist of (A, a) or (A, b) or both. From Theorem 3, it follows that in 2×2 games, a strictly dominated strategy will never be used with positive probability in the long run. However, as illustrated by the game in Figure 1.10, a weakly dominated strategy may be used with positive probability. In this game, the unique minimal construction is $V = \{(A, b), (B, b), (B, a)\}$.

	a	b
A	0,0	3,3
B	3,3	3,3

FIGURE 1.10

Moreover, in games with more than two strategies, strictly dominated strategies may be played with positive probability. Consider the game in Figure 1.11.

	a	b
A	0,3	2,0
B	2,0	1,3
C	1,1	0,1

FIGURE 1.11

In this game, B strictly dominates C, but the unique minimal construction is $V = \{(A,a), (C,a), (C,b), (B,b), (A,b), (B,a)\}$ and by Theorem 3, imitation play will put positive probability on all pure-strategy profiles in this minimal construction in the long run. Play of strictly dominated strategies may appear an undesirable or counter intuitive feature, but it is merely a consequence of the limited information available to the individuals in the model. Since they do not know the true payoffs to every pure-strategy profile, but rely on information provided by a finite sample of previous play, they cannot deduce that a certain strategy is dominated.

1.4.4 Payoff Transformations

Minimal constructions are invariant under affine payoff transformations and local payoff shifts, which follows since the better-reply graph is unaffected by such changes. The stochastically stable states of imitation play are also invariant under affine payoff transformations, but not under local payoff shifts. The two games in Figure 1.12 illustrate the latter claim.

	a	b
A	3,3	0,0
B	0,0	2,2

	a	b
A	3,3	2,0
B	0,2	4,4

FIGURE 1.12

In the left game, (A, a) is stochastically stable according to Theorem 4. By adding 2 units to the each of the row player's payoffs in column b and to each of the column player's payoff in row B in this game, the right game is obtained. In the right game, (B, b) is stochastically stable by Theorem 4.

1.5 Applications to 2×2 Games

In this section, we apply the results from Section 3 to 2×2 games. First, we find the stochastically stable states in three special classes of games and second, we study the speed of convergence in general 2×2 games. In all of the following games, we denote player 1's strategies A and B, and player 2's strategies a and b, respectively.

1.5.1 Stochastically Stable States in Special Classes of Games

In this subsection, we analyze the stochastic stability in games with two strict Nash equilibria, games with one strict Nash equilibrium and games without Nash equilibria in pure strategies. We start with the class of games with a unique strict Nash equilibrium, which includes, for example, *Prisoners' Dilemma Games.*

Proposition 2 *In 2×2 games with a unique strict Nash equilibrium, if $s/m \leq 1/2$, the corresponding monomorphic state is a unique stochastically stable state.*

PROOF: Games in this class contain exactly one minimal construction, consisting of the unique strict Nash equilibrium. By Theorem 3, the corresponding monomorphic state must be the unique stochastically stable state. *Q.E.D.*

We now proceed with the class of games with two strict Nash equilibria. *Coordination Games* and *Hawk-Dove Games* are two examples of games in this class.

Proposition 3 *In 2×2 games with two strict Nash equilibria, where one Nash Equilibrium strictly Pareto dominates the other, if $s/m \leq 1/2$ and s is sufficiently large, then the unique stochastically stable state corresponds one to one with the monomorphic state of the Pareto dominant equilibrium.*

PROOF: Games in this class contain two minimal constructions, either $\{(A,a)\}$ and $\{(B,b)\}$ or $\{(A,b)\}$ and $\{(B,a)\}$. Without loss of generality, assume that $\{(A,a)\}$ and $\{(B,b)\}$ are the minimal constructions. If (A,a) strictly Pareto dominates (B,b), it must also strictly Pareto dominate the two other pure-strategy profiles, (A,b) and (B,a). Hence, by Corollary 1, the monomorphic state $h_{(A,a)}$ is the unique stochastically stable state for $s/m \leq 1/2$ and s is sufficiently large. *Q.E.D.*

This implies that, unlike Young's (1993) process of adaptive play, imitation play does not generally converge to the risk-dominant equilibrium in Coordination Games. In the right game in Figure 1.12, (A,a) is a risk dominant equilibrium whereas (B,b) is a strictly Pareto superior equilibrium and subsequently, the unique stochastically stable state of imitation play. Our result is consistent with Robson and Vega-Redondo's (1996) result for generic symmetric Coordination Games. However, for the non-generic case when $\pi_i(A,a) = \pi_i(B,b)$, the stochastically stable states in their model depend on the details of the adjustment process, whereas imitation play always selects both equilibria. Consider the following proposition:

Proposition 4 *In 2×2 games with two strict Nash equilibria, where neither strictly Pareto dominates the other, if $s/m \leq 1/2$ and s is sufficiently large, then the stochastically stable states correspond one to one with the monomorphic states of the two equilibria.*

PROOF: See the Appendix.

Finally, we consider games which do not have any Nash equilibrium in pure strategies. One of the games in this class is the *Matching Pennies Game.*

Proposition 5 *In 2×2 games without Nash equilibria in pure strategies, if $s/m \leq 1/2$, the stochastically stable states correspond one to one with the four monomorphic states of the game.*

PROOF: Games in this class contain exactly one minimal construction $\{(A,a),(A,b),(B,a),(B,b)\}$. By Theorem 3, the four corresponding monomorphic states must all be stochastically stable. *Q.E.D.*

1.5.2 Speed of Convergence

In this subsection, we analyze the speed of convergence of imitation play.

Proposition 6 *In 2×2 games, if $s/m \leq 1/2$, the maximum expected first passage time for the perturbed process $P^{\alpha,m,s,\varepsilon}$ from any state to a stochastically stable state is at most $\delta\varepsilon^{-2}$ units of time, for some positive constant δ.*

PROOF: The claim follows from the observation that in any 2×2 game, the transition from an arbitrary state to the basin of attraction of a stochastically stable monomorphic state requires two mistakes at most. *Q.E.D.*

This result should be contrasted with the speed of convergence in Young (1993), Kandori *et al.* (1993), and Robson and Vega-Redondo (1996). In Young's (1993) model, the maximum expected first passage for a 2×2 Coordination Game is at least $\delta_Y\varepsilon^{-v}$ where v depends on the sample size and both players' payoffs. In Kandori *et al.* (1993) the maximum expected first passage time is of the order $\delta_{KMR}\varepsilon^{-Nu}$, where N is the size of the population and u is determined by the game's payoff structure. In Robson and Vega-Redondo (1996), the corresponding figure is $\delta_{RV}\varepsilon^{-q}$, where q is a positive integer independent of the payoffs and the current state. Thus, when v, Nu, and q are greater than two and ε is sufficiently small, then imitation play converges considerably faster than the processes in these three models.

1.6 Extensions

All results in this paper hold for a more general class of imitation dynamics. Let the *maximum correspondence* be a correspondence which maps a strategy sample w_i and the associated payoff sample v_i to a strategy with the highest payoff in the sample. This correspondence defines a new Markov process on the space of histories with the same set of absorbing states and (for a sufficiently large sample size) stochastically stable states as imitation play. Moreover, if each population consists of arbitrary shares of individuals who make choices based on the maximum correspondence and the maximum average correspondence, respectively, then the results of this paper still hold. Hence, the model allows for a certain kind of population heterogeneity, where individuals make their choices based on different rules.

1.7 Conclusion

In this paper we develop an evolutionary model with perpetual random shocks where individuals, in every period, choose the strategy with the highest average payoff in a finite sample of past play. We denote the resulting

Markov process imitation play and prove that, provided information is sufficiently incomplete and the sample size sufficiently large, the stochastically stable states of imitation play are repetitions of the pure-strategy profiles in minimal closed sets under the better-reply graph. We call such sets minimal constructions. These sets are related to minimal closed sets under better replies and other set-wise concepts. We also prove that if the pure-strategy profiles in a minimal construction have strictly higher payoffs than all outside pure-strategy profiles, then, provided that the minimal construction is a product set, the pure-strategy profiles in this set will be played with probability one in the limit as the experimentation probability tends to zero. Our results give clear predictions in 2×2 games. In Coordination Games, where one equilibrium strictly Pareto dominates the other, imitation play selects the strictly Pareto superior Nash equilibrium. If neither equilibria strictly Pareto dominates the other, then both are stochastically stable. Finally, we show that the speed of convergence for imitation play in many cases is higher than in other known models.

The objective of this paper is to derive predictions for general finite games in a world of truly boundedly rational individuals. The assumption underlying the model, that individuals do not make decisions based on the predictions of their opponents' future strategies, but rather based on which strategies have been successful in the past, is maybe most appealing in the class of games where it is costly to obtain information about the opponents. A high cost may be due to the size or the complexity of the game or to institutional factors preventing the release of information about the opponents. It would be particularly interesting to test the implications of our model against empirical or experimental evidence in this class of games.

1.8 Appendix

PROOF OF LEMMA 1: This proof follows Young (1998). Given a history $h^t = (x^{t-m+1}, ..., x^t)$ at time t, the process moves to a state of the form $h^{t+1} = (x^{t-m+2}, ..., x^t, x^{t+1})$ in the next period. Remember that such a state is called a *successor* of h^t. The strategy x_i^{t+1} is an *idiosyncratic choice* or *error* if and only if there exists no sample of strategy choices w_i and payoff outcomes v_i, such that $x_i^{t+1} \in \alpha_i(w_i, v_i)$. For each successor h^{t+1} of h^t, let $r_{h^t h^{t+1}}$ denote the total number of errors in the transition from h^t to h^{t+1}. Evidently, $0 \leq r_{h^t h^{t+1}} \leq n$ and the probability of the transition $h^t \to h^{t+1}$ is in the order of $\varepsilon^{r_{h^t h^{t+1}}}(1-\varepsilon)^{n-r_{h^t h^{t+1}}}$. If h^{t+1} is not a successor of h^t, the probability of the transition is zero. Hence, the process $P^{\alpha,m,s,\varepsilon}$ approaches $P^{\alpha,m,s,0}$ at an exponential rate in ε. Since the process is also irreducible whenever $\varepsilon > 0$, it follows that $P^{\alpha,m,s,\varepsilon}$ is a regular perturbed Markov process. *Q.E.D.*

PROOF OF THEOREM 3: In order to prove the claim, we will show that:
A) all monomorphic states in a minimal construction have equal stochastic potential;
B) for every monomorphic state which does not belong to any minimal construction, there exists a monomorphic state with lower stochastic potential. The theorem follows immediately from A) and B), by applying Lemma 1 and Theorem 2.

Let us start from A). On the one hand, by Lemma 3, there exists a better-reply path from an arbitrarily monomorphic state h_x in a minimal construction to any other monomorphic state $h_{x'}$ in the same minimal construction. Let the sequence $h_x, ..., h_{x'}$ be such a path. By Lemma 2, the following inequalities hold:

$$\rho(x) \geq ... \geq \rho(x'). \tag{1.6}$$

On the other hand, by applying Lemma 3 once more, there exist a better-reply path from monomorphic state $h_{x'}$ to monomorphic state h_x. Using Lemma 2, gives:

$$\rho(x') \geq ... \geq \rho(x). \tag{1.7}$$

From the inequalities in (1.6) and (1.7) it follows that $\rho(x) = \rho(x')$ for any monomorphic states h_x and $h_{x'}$ in a minimal construction.

Part B). Note that for every monomorphic state not included in any minimal construction, there exists a finite better-reply path which ending in some minimal construction. Let this path be $h_{x^1}, h_{x^2}, ..., h_{x^{T-1}}, h_{x^T}$, where h_{x^1} is an arbitrary monomorphic state that does not belong to any minimal construction and h_{x^T} the first monomorphic state on the path belonging to some minimal construction, V. By Lemma 2, it follows that

$$\rho\left(x^1\right) \geq \rho\left(x^2\right) \geq ... \geq \rho\left(x^{T-1}\right) \geq \rho\left(x^T\right). \tag{1.8}$$

We will show that in fact, $\rho\left(x^{T-1}\right) > \rho\left(x^T\right)$. Note that $\rho\left(x^{T-1}\right)$ is the minimum resistance over all trees rooted at the state $h_{x^{T-1}}$. Denote (one of) the tree(s) that minimizes resistance by $\Im\left(x^{T-1}\right)$. Find in the tree $\Im\left(x^{T-1}\right)$ a directed edge from some vertex $y \in X$ in the minimal construction V to some other vertex $y' \in X$ outside this minimal construction. It will be shown later that there is only one such directed edge in the minimal resistance tree $\Im\left(x^{T-1}\right)$. Delete in the tree $\Im\left(x^{T-1}\right)$ the directed edge $y \to y'$ and add the directed edge $x^{T-1} \to x^T$. As a result, we obtain a tree $\Im(y)$ rooted at the state h_y. By construction, the total resistance of the new tree $\Im(y)$ is less than the stochastic potential $\rho\left(x^{T-1}\right)$. Moreover, by part A), the monomorphic state h_{x^T} of the minimal construction has the same stochastic potential as the monomorphic state h_y. Hence, $\rho\left(x^{T-1}\right) > \rho\left(x^T\right)$.

We will now consider the tree $\Im\left(x^{T-1}\right)$ and show that there is only one directed edge from every minimal construction to a state outside the construction. Suppose there is a finite number of such directed edges $y^j \to z^j$, $j = 1, 2, ..., l$ from some minimal construction, where $y^1, ..., y^l$ are vertices in the minimal construction and $z^1, ..., z^l$ vertices outside the construction. It is clear that there cannot be an infinite number of outgoing edges since the game Γ is finite. Recall that a tree rooted at vertex y^j is a set of $|X| - 1$ directed edges such that, from every vertex different from y^j, there is a unique directed path in the tree to y^j. The resistance of any directed edge $y^j \to z^j$, $j = 1, 2, ..., l$ is at least two. By Lemma 3, there exists a finite better-reply path from vertex y^1 to vertex y^2 in the minimal construction. Let $y^1, f^1, ..., f^k, y^2$ be such a path.

Consider vertex f^1. There are two mutually exclusive cases:
1.a) there exists a directed path from f^1 to one of the vertices $y^2, ..., y^l$ in the initial tree $\Im\left(x^{T-1}\right)$, or
1.b) there exists a directed path from f^1 to y^1.

In case 1.a) by deleting the directed edge $y^1 \to z^1$ and adding the directed edge $y^1 \to f^1$ to the tree $\Im\left(x^{T-1}\right)$, we obtain a new tree $\Im^1\left(x^{T-1}\right)$ with lower stochastic potential than $\Im\left(x^{T-1}\right)$, because the resistance of the directed edge $y^1 \to f^1$ is one. This means that we are done, since this contradicts the assumption of $\Im\left(x^{T-1}\right)$ being a minimal resistance tree.

In case 1.b), we will use the following procedure for vertex f^1: delete the initial directed edge from f^1 and add the directed edge $f^1 \to f^2$. As above, there are two cases:

2.a) there exists a directed path from f^2 to one of the vertices $y^2, ..., y^l$ in the initial tree $\Im\left(x^{T-1}\right)$, or
2.b) there exists a directed path from f^2 to y^1.

In case 2.a), we obtain a new tree $\Im^2\left(x^{T-1}\right)$ with lower stochastic potential than $\Im\left(x^{T-1}\right)$, because the resistance of the directed edge $f^1 \to f^2$ is one. This means that we are done, since it contradicts the assumption of $\Im\left(x^{T-1}\right)$ being a minimal resistance tree.

In case 2.b), we repeat the procedure for vertices $f^2, f^3, ...$ The better-reply path $y^1, f^1, ..., f^k, y^2$ from vertex y^1 to vertex y^2 is finite. Hence, after $k+1$ steps at most, we have constructed a tree $\Im''\left(x^{T-1}\right)$ rooted at the state $h_{x^{T-1}}$ with lower stochastic potential than $\Im\left(x^{T-1}\right)$, which is impossible. *Q.E.D.*

PROOF OF THEOREM 4: Define the *basin of attraction* of a state h' as the set of states h such that there is a positive probability of moving from h to h' in a finite number of periods under the unperturbed process $P^{\alpha,m,s,0}$. Let V be a minimal construction which strictly Pareto dominates all pure-strategy profiles outside V. Let h_y be a monomorphic state such that $y \notin V$. Consider a minimal resistance tree $\Im(y)$ rooted at h_y. Let $x \in V$ be a vertex such that there is a directed edge from x to a vertex $z \notin V$ in the tree $\Im(y)$. Assume the stochastic potential of h_x to be at least as high as the stochastic potential of h_y.

We claim that for a sufficiently large s, the resistance of the edge from $x \in V$ to $z \notin V$ is greater than n. Without loss of generality, assume that $s > n$ (otherwise s consecutive mistakes by individuals in one population could be sufficient to transfer from h_x to the basin of attraction of h_z). Let $\pi_i(x)$ and $\pi_i(z)$ be payoffs for player i corresponding to the pure-strategy profiles x and z respectively. The individual in player role i will play z_i instead of x_i only if the average payoff for z_i in her sample is greater than or equal to the average payoff for x_i. With n mistakes at most, this is only possible if the following inequality holds

$$\frac{(s-n)\pi_i(x) + (n-1)\pi_i(x_i, x'_{-i})}{s-1} \leq \pi_i(z), \tag{1.9}$$

where $\pi_i(x_i, x'_{-i})$ is the payoff to player i when she plays x_i and her opponents play any $x'_{-i} \neq x_{-i}$. It is clear that this inequality does not hold for any individual in player role i, if s is sufficiently large. Hence, the resistance from h_x to h_z must be larger than n for a sufficiently large s.

Create a tree $\Im(x)$ rooted at h_x by adding a directed edge from h_y to h_x and deleting the directed edge from x to z in the tree $\Im(y)$. Provided that s is sufficiently large, the deleted edge has a resistance greater than n

and the added edge has a resistance of at most n (n simultaneous mistakes are sufficient to move from any monomorphic state to the basin of attraction of h_x). Hence, the total resistance of the new tree $\Im(x)$ is less than $\rho(y)$, contradicting the assumption of $\rho(y) \leq \rho(x)$. This establishes that h_x has minimal stochastic potential. Theorem 4 now follows from part A) in the proof of Theorem 3, where it is proved that all monomorphic states of a minimal construction have equal stochastic potential. *Q.E.D.*

PROOF OF PROPOSITION 4: As in the proof of Proposition 2, without loss of generality, assume that $\{(A, a)\}$ and $\{(B, b)\}$ are the minimal constructions of the game. By Theorem 3, it follows that monomorphic states $h_{(A,a)}$ and $h_{(B,b)}$ are the only two candidates for the stochastically stable states. Suppose that only one of these monomorphic states is stochastically stable, say $h_{(A,a)}$. Let $\Im(A, a)$ be a minimum resistance tree with resistance $\rho(A, a)$ rooted at (A, a). In this tree, there is an outgoing edge from the monomorphic state $h_{(B,b)}$.

First, note that the resistance of this edge is at least two, such that at least two mistakes are needed to move from the monomorphic state $h_{(B,b)}$. This follows since $\pi_1(B, b) > \pi_1(A, b)$ and $\pi_2(B, b) > \pi_2(B, a)$.

Second, note that provided that $s/m \leq 1/2$ and s is sufficiently large, two mistakes are sufficient to move the process from monomorphic state $h_{(A,a)}$ to monomorphic state $h_{(B,b)}$. Suppose the process is in state $h_{(A,a)}$. Since neither of the Nash equilibria is strictly Pareto superior, either $\pi_1(B, b) \geq \pi_1(A, a)$ and/or $\pi_2(B, b) \geq \pi_2(A, a)$. Without loss of generality, assume that the first of these inequalities holds. Then, there is a positive probability that the individuals in both player roles simultaneously make mistakes at time t. There is also a positive probability that the individuals in player role 1 draw the fixed sample $(x_1^{t-s+1}, ..., x_1^t)$ with corresponding payoffs $(\pi_1^{t-s+1}, ..., \pi_1^t)$ and that the individuals in player role 2 sample from plays earlier than x_2^t from period $t+1$ to, and including, period $t+s-1$. With positive probability individuals in player role 1 play B and individuals in player role 2 play a in all of these periods. This implies that if the individuals in both player roles sample from plays more recent than x_i^{t-1} from period $t + s$ to, and including, period $t + s - 1 + m$, 1's sample will only contain strategy B and 2's sample will always contain strategy b, and possibly, strategy a. Furthermore, the average payoff of strategy b will be $\pi_2(B, b)$ as compared to an average payoff of $\pi_2(B, a)$ for strategy a, when the latter strategy is included in the sample. Hence, with positive probability the process will be in state $h_{(B,b)}$ at time $t + s - 1 + m$.

Finally, create a new tree rooted at $h_{(B,b)}$ by deleting the outgoing edge from the monomorphic state $h_{(B,b)}$ in the tree $\Im(A, a)$ and adding an edge from $h_{(A,a)}$ to $h_{(B,b)}$. The resistance of the deleted edge is at least two and that of the added edge two. Hence, the total resistance of the new tree is at most $\rho(A,a)$, thereby contradicting the assumption that only $h_{(A,a)}$ is stochastically stable. *Q.E.D.*

Bibliography

[1] Basu, K., and Weibull J. (1991). "Strategy Subsets Closed under Rational Behavior," *Economics Letters* 36, 141-146.

[2] Björnerstedt, J., and Weibull J. (1996). "Nash Equilibrium and Evolution by Imitation," in *The Rational Foundations of Economic Behaviour* (K. Arrow et al. Eds.), 155-171. MacMillan.

[3] Duffy, J., and Feltovich N. (1999). "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study," *International Journal of Game Theory* 28, 131-152.

[4] Graham, J. R. (1999). "Herding among Investment Newsletters: Theory and Evidence," *Journal of Finance* 54, 237-268.

[5] Griffiths, M. D., Smith, B. F., Turnbull, D., Alasdair, S., and White R. W. (1998). "Information flows and open outcry: evidence of imitation trading," *Journal of International Markets, Institutions & Money* 8, 101-116.

[6] Huck, S., Normann, H.-T., and Oechssler J. (1999). "Learning in Cournout Oligopoly: An experiment," *Economic Journal* 109, 80-95.

[7] ——— (2000). "Does information about competitors' actions increase or decrease competition in experimental oligopoly markets?," *International Journal of Industrial Organisation* 18, 39-58.

[8] Foster, D., and P. Young (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* 38: 219-32.

[9] Freidlin, M., and A. Wentzell (1984): *Random Perturbations of Dynamical Systems.* Springer-Verlag.

[10] Kandori, M., Mailath, G., and Rob R. (1993). "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica* 61, 29-56.

[11] Nash, J. (1950). "Non Cooperative Games," Ph.D. thesis, Princeton University.

[12] Nöldeke, G., and Samuelson L. (1993). "An Evolutionary Analysis of Backward and Forward Induction", *Games and Economic Behavior* 5, 425-454.

[13] Osborne, M. J., and Rubinstein A. (1998). "Procedurally Rational Players," *American Economic Review* 88, 834-847.

[14] Ritzberger, K., and Weibull J. (1995). "Evolutionary Selection in Normal-Form Games" *Econometrica* 63, 1371-1400.

[15] Robson, A. J., and Vega-Redondo F. (1996). "Efficient Equilibrium Selection in Evolutionary Games with Random Matching", *Journal of Economic Theory* 70, 65-92.

[16] Schlag, K. (1998). "Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-armed Bandits," *Journal of Economic Theory* 78, 130-56.

[17] —— (1999). "Which One Should I Imitate?," *Journal of Mathematical Economics* 31, 493-522.

[18] Sobel, J. (1993). "Evolutionary Stability and Efficiency," *Economic Letters* 42, 301-312.

[19] Wermers, R. (1999). "Mutual Fund Herding and the Impact on Stock Prices,", *Journal of Finance* 54, 237-268.

[20] Young, P. H. (1993). "The Evolution of Conventions," *Econometrica* 61, 57-84.

[21] —— (1998). *Individual Strategy and Social Structure.* Princeton University Press.

Chapter 2

Stochastic Better-Reply Dynamics in Finite Games

2.1 Introduction

In Young (1993, 1998), individuals from each of n finite populations are randomly selected to play a finite n-player game every period. Each of the individuals forms beliefs of her opponents' play by inspecting an independent sample of recent strategy realizations from a finite history of past play and calculating the corresponding frequency distributions over her opponents' pure strategies. Then, she chooses a pure strategy which is a best reply given her beliefs. With a small probability, she instead experiments or makes a mistake and play any pure strategy at random. This results in an ergodic Markov chain with a unique stationary distribution on the space of histories.

In the present paper I alter Young's behavioral assumption by allowing individuals to play not only best replies but also *better replies,* given their beliefs. Formally, the *better-reply correspondence* maps distributions over the player's own and her opponents' pure strategies to those pure strategies which give the player at least the same expected payoff against the distribution of her opponents' pure strategies. This behavior could interpreted as a special case of Simon's (1955) *satisfying* behavior, with the aspiration level given by the expected payoff to the sample distribution. The best-reply and the better-reply correspondences have important properties in common, among other things, they both belong to the class of *behavior correspondences* (Ritzberger and Weibull, 1995). This class contains upper hemi-continuous correspondences such that the image of any product distribution includes the set of best-replies. The similarities between the two correspondences imply that Young's (1998) predictions in weakly acyclic games and in 2×2 Coordination

Games are unaffected if best replies are replaced by better replies.

A product set of pure strategies is *closed* under the better-reply correspondence if, under this correspondence, the image of every distribution with support in the set is contained in the set. It is a *minimal closed set* if it does not properly contain a subset with this property. Such sets are important in my analysis of the stochastic adaptive model where individuals apply the better-reply correspondence. I show that in finite n-player games, as the experimentation probability tends to zero, only strategies in certain minimal closed sets under better replies are played with positive probability. This is analogous to the result in Young (1998) for the best-reply model. He proves that in a generic class of finite games, only strategies in certain minimal closed sets under best replies will be played with positive probability in the limit. Due to a specific property of the better-reply correspondence, my result is valid for all games, whereas Young has to assume the game is nondegenerate in a specific sense.

My result is also consistent with and extends previous findings on deterministic dynamics in a continuous-time setting. Ritzberger and Weibull (1995) prove equivalence between asymptotically stable sets under sign-preserving selection dynamics, and closed sets under better-replies. Hence, my model provides a way of selecting among such asymptotically stable sets.

This paper is organized as follows. In Section 2, I present the stochastic process and introduce basic definitions. Section 3 contains general results on the convergence of the stochastic process, and section 4 results for two special classes of games. In Section 5, I relate my findings to previous results on set-wise stability under regular selection dynamics. Section 6 contains examples and Section 7 discusses extensions of the model. The proof of the main theorem can be found in the Appendix.

2.2 Model

The basic setting is similar to that of Young (1993, 1998). Let Γ be a finite n-player game in strategic form. Let X_i be the finite set of pure strategies x_i available to player $i \in \{1, ..., n\} = N$ and let $\Delta(X_i)$ be the set of probability distributions p_i over these strategies. Define the product sets $X = \prod X_i$ and $\square(X) = \prod_i \Delta(X_i)$, with the typical elements x and p respectively. Let $p_i(x_i)$ denote the probability mass on pure strategy x_i and let $p(x) = \prod_{i \in N} p_i(x_i)$. The notation $x_{-i} \in \prod_{j \neq i} X_j$ and analogously, $p_{-i} \in \prod_{j \neq i} \Delta(X_j) = \square(X_{-i})$, is used to represent the strategies and distributions of strategies of players other than i. Let $C_1, ..., C_n$ be n finite and non-empty populations of individuals. Each member of population C_i is a candidate to play role i in the game

G and has payoffs represented by the utility function $\pi_i : X \to \mathbb{R}$, and expected payoffs represented by the function $u_i : \square(X) \to \mathbb{R}$. In slight abuse of notation, I write $u_i(x_i, p_{-i})$ instead of $u_i(p_i, p_{-i})$ if $p_i(x_i) = 1$.

The following two correspondences are instrumental in the subsequent analysis.

Definition 1 *Let $\beta = \prod_i \beta_i : \square(X_{-i}) \twoheadrightarrow X$ be the best-reply correspondence, defined by $\beta_i(p_{-i}) = \{x_i \in X_i \mid u_i(x_i, p_{-i}) - u_i(x_i', p_{-i}) \geq 0 \ \forall x_i' \in X_i\}$.*

Definition 2 *(Ritzberger and Weibull, 1995) Let $\gamma = \prod \gamma_i : \square(X) \twoheadrightarrow X$ be the better-reply correspondence, defined by $\gamma_i(p) = \{x_i \in X_i \mid u_i(x_i, p_{-i}) - u_i(p) \geq 0\}$.*

In other words, γ_i assigns to each product distribution $p \in \square(X)$ those pure strategies x_i which give i at least the same expected payoff as p_i.[1] Note that the better-reply correspondence is defined as a mapping from the product of the simplices for *all* players, whereas the best-reply correspondence is a mapping from the product of only the opponents' simplices.[2]

Let $t = 1, 2, \ldots$ denote successive time periods. The stage game Γ is played once each period. In period t, one individual is drawn at random from each of the n populations and assigned to play the corresponding role. The individual in role i chooses a pure strategy x_i^t from her strategy space X_i according to a rule that will be defined below. The pure-strategy profile $x^t = (x_1^t, \ldots, x_n^t)$ is recorded and referred to as *play* at time t. The *history* or *state* at time $t \geq m$ is the sequence $h^t = (x^{t-m+1}, \ldots, x^t)$, where m denotes the *memory size* of all individuals. Let $H = X^m$ be the finite set of histories of length m and let h be an arbitrary element of this set.

Strategies are chosen as follows. Assume an arbitrary initial history $h^m = (x^1, \ldots, x^m)$ at time m. In every subsequent period, each individual drawn to play the game inspects s plays, drawn without replacement, from the most recent m periods. The draws are independent for the various individuals and across time. For each $x_i \in X_i$, let $p_i(x_i \mid h)$ be the conditional probability that the individual in role i chooses pure strategy x_i, given history h. I assume that $p_i(x_i \mid h)$ is independent of t and that $p_i(x_i \mid h) > 0$ if and only if there exists a sample of size s from the history h, consisting of n independent draws and with a sample distribution of $\hat{p} \in \square(X)$, such that

[1] It is evident that γ_i is u.h.c. and $\beta_i(p) \subset \gamma_i(p)$ for all players and product distributions. As will be discussed below, this implies that γ is a *behavior correspondence* (Ritzberger and Weibull, 1995).

[2] The best-reply correspondence can, naturally, also be represented as a mapping from the set of product distributions to the set of pure strategy-tuples $\beta = \prod_i \beta_i : \square(X) \twoheadrightarrow X$, with $\beta_i(p) = \beta_i(p_i', p_{-i})$ for all $p_i' \in \Delta(X_i)$.

$x_i \in \gamma_i(\hat{p})$, where γ_i is the better-reply correspondence defined above. Unlike in Young (1993, 1998), the individuals here play a better reply, and not a best reply to this sample distribution.

Given a history $h^t = (x^{t-m+1}, ..., x^t)$ at time t, the process moves to a state of the form $h^{t+1} = (x^{t-m+2}, ..., x^t, x^{t+1})$ in the next period. Such a state is called a *successor* of h^t. My behavioral assumptions imply that the process moves from a current state h to a successor state h' in each period according to the following transition rule. If x is the rightmost element of h, the probability of moving from h to h' is $P^{\gamma,m,s,0}_{hh'} = \prod_{i=1}^{n} p_i(x_i \mid h)$ if h' is a successor of h and 0 if it is not. This defines a finite Markov chain on the finite state space of histories H. I call the process $P^{\gamma,m,s,0}$ *γ-adaptive play with memory m and sample size s* and I generally refer to it as the unperturbed process. A *recurrent class* E_k of the process $P^{\gamma,m,s,0}$ is a set of states such that there is zero probability of moving from any state in the class to any state outside, and there is a positive probability of moving from any state in the class to any other state in the class. A state h is *absorbing* if it constitutes a singleton recurrent class.

I also define a perturbed process in a manner completely analogous to Young (1993, 1998). Formally, in each period, there is some small probability ε, that the individual in role i experiments by choosing a pure strategy at random from X_i, instead of according to the better-reply correspondence. The event that the individual in role i experiments is assumed to be independent of the event that the individual in role j experiments for every $j \neq i$. For every i, let $q_i(x_i \mid h)$ be the conditional probability that the individual in role i chooses $x_i \in X_i$, given that she experiments and the process is in state h. I assume that $q_i(x_i \mid h)$ is independent of t and that $q_i(x_i \mid h) > 0$ for all $x_i \in X_i$. Suppose that the process is in state h at time t. Let J be a subset of j player roles. The probability is $\varepsilon^j(1-\varepsilon)^{n-j}$ that exactly the individuals in player roles of J experiment and the others do not. Conditional on this event, the transition probability of moving from h to h' is $Q^J_{hh'} = \prod_{i \in J} q_i(x_i \mid h) \prod_{i \notin J} p_i(x_i \mid h)$ if h' is a successor of h and x is the rightmost element of h' and 0 if h' is not a successor of h. This gives the following transition probability of the perturbed Markov process:

$$P^{\gamma,m,s,\varepsilon}_{hh'} = (1-\varepsilon)^n P^{\gamma,m,s,0}_{hh'} + \sum_{J \subset N, J \neq \emptyset} \varepsilon^{|J|}(1-\varepsilon)^{n-|J|} Q^J_{hh'}.$$

I call the process $P^{\gamma,m,s,\varepsilon}$ *γ-adaptive play with memory m, sample size s, experimentation probability ε and experimentation distributions q_i.*

This process is irreducible and aperiodic, and thus has a unique stationary distribution μ^ε, which I study as ε tends to zero. In my analysis, I use the following standard definitions, due to Freidlin and Wentzell (1984), and

Foster and Young (1990). A state h is *stochastically stable* if $\lim_{\varepsilon \to 0} \mu^{\varepsilon}(h)$ exists and is positive. For each pair of distinct recurrent classes E_k and E_l, a *kl-path* is a sequence of states $\zeta = (h^1, h^2 ..., h^q)$ that begins in E_k and ends in E_l. The *resistance* of this path is the sum of the resistances on the edges that compose it. Let r_{kl} be the least resistance over all kl-paths. Construct a complete directed graph with one vertex for each recurrent class. The weights on the directed edge $E_k \longrightarrow E_l$ is r_{kl}. A *tree* rooted at E_l is a set of directed edges such that, from every vertex different from E_l, there is a unique directed path in the tree to E_l. The *resistance* of a rooted tree is the sum of the resistances on the edges that compose it. The *stochastic potential* $\rho(E_l)$ of a recurrent class E_l is the minimum resistance over all trees rooted at E_l.

2.3 General Results

In this section, I investigate the properties of the limiting stationary distribution of γ-adaptive play as the experimentation probability tends to zero.[3] In what follows, I will refer to the stochastic process in Young (1993, 1998) as β-adaptive play in order to distinguish it from γ-adaptive play. I will first prove my main result that in finite n-player games, the limiting distribution of γ-adaptive play puts positive probability only on histories with support in certain minimal closed sets under better replies. This result is similar to Theorem 7.2 in Young (1998) for β-adaptive play, but holds more generally.

Some further definitions are needed to state my main result. Let $\mathcal{X}$ be the collection of all nonempty product sets $Y \subset X$. Let $\Delta(Y_i)$ be the set of probability distribution with support in Y_i and let $\square(Y) = \prod_{i \in N} \Delta(Y_i)$ be the corresponding product set.

Definition 3 *(Ritzberger and Weibull, 1995) A set $Y \in X$ is* ***closed under better replies (CUBR)*** *if $\gamma(\square(Y)) \subseteq Y$. A set $Y \in X$ is a* ***minimal closed set under better replies (MCUBR)*** *if it is closed under better replies and contains no proper subset with this property.*

Important to my analysis is the following lemma, which implies that my main theorem applies to *any* finite game and not only to a generic class of games as the corresponding theorem in Young (1998). Define $\tilde{\gamma}$ as the better-reply correspondence with pure strategy domain. That is, $\tilde{\gamma} = \prod \tilde{\gamma}_i : X \twoheadrightarrow X$, where
$\tilde{\gamma}_i(x) = \{x'_i \in X_i \mid \pi_i(x'_i, x_{-i}) - \pi_i(x) \geq 0\}$.

[3] This limiting distribution exists by Lemma 2 and Theorem 5 in the Appendix.

Lemma 1 *If $x_i \in \gamma_i(p)$, then there exists a pure-strategy profile $y \in X$ with $p(y) > 0$, such that $x_i \in \tilde{\gamma}_i(y)$.*

PROOF: This proof uses the multilinearity of the function u_i. Consider an arbitrary distribution $p \in \square(X)$, and a pure strategy $x_i \in X_i$, such that $x_i \in \gamma_i(p)$. I will first show that $\exists y_i \in X_i$ with $p_i(y_i) > 0$, such that $u_i(x_i, p_{-i}) - u_i(y_i, p_{-i}) \geq 0$. By definition of the better-reply correspondence, $u_i(x_i, p_{-i}) - u_i(p) \geq 0$. Write this difference in the following way:

$$u_i(x_i, p_{-i}) - u_i(p) = \sum_{y_i \in X_i} p_i(y_i) \left[u_i(x_i, p_{-i}) - u_i(y_i, p_{-i})\right]. \qquad (2.1)$$

Clearly, if the left-hand side is non-negative, at least one of the terms in the sum on the right-hand side with $p_i(y_i) > 0$ must be non-negative. Hence, $\exists y_i \in X_i$ with $p_i(y_i) > 0$, such that $u_i(x_i, p_{-i}) - u_i(y_i, p_{-i}) \geq 0$.

Second, write the last difference in the following way:

$$u_i(x_i, p_{-i}) - u_i(y_i, p_{-i}) = \sum_{y_{-i} \in X_{-i}} p_{-i}(y_{-i}) \left[\pi_i(x_i, y_{-i}) - \pi_i(y_i, y_{-i})\right]. \qquad (2.2)$$

By the same logic as above, if the left-hand side is non-negative, at least one of the elements in the sum on the right-hand side with $p_{-i}(y_{-i}) > 0$ must be non-negative. Hence, $\exists y \in X$ with $p(y) > 0$, such that $\pi_i(x_i, y_{-i}) - \pi_i(y) \geq 0$ *Q.E.D.*

Lemma 1 says that if a pure strategy x_i is a better reply to the product distribution p, then there also exists a pure-strategy profile in the support of p, to which x_i is a better reply. This implies that the set of better replies to all product distributions over a particular product set of pure-strategy profiles is identical to the set of better replies to the product set of pure-strategy profiles.

Corollary 1 $\gamma(\square(Y)) = \tilde{\gamma}(Y)$ *for all* $Y \in \mathcal{X}$.

It is worth noting that a result analogous to Lemma 1 does not hold for the best-reply correspondence. Consider the game in Figure 2.1, and the distribution $p^* = ((\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, \frac{1}{2}))$ with the associated expected payoff $u_1(p^*) = \frac{3}{2}$ to player 1. Clearly, C is a best reply for player 1 to p_2^* and hence, also a better reply to p^*. Moreover, there exists a pure-strategy profile (B, a) in the support of p^* to which C is a better reply. However, there exists no pure strategy in the support of p_2^* to which C is a best reply, since $\beta_1^{-1}(C) = \{p_2 \in \Delta(X_2) \mid \frac{1}{3} \leq p_2(a) \leq \frac{2}{3}\}$.

	a	b
A	$3,1$	$0,0$
B	$0,0$	$3,1$
C	$2,0$	$2,0$

FIGURE 2.1

Recall that $H = X^m$ is the space of histories. Let H' be a subset of this set and let $S(H')$ be the *span* of this subset, i.e. the product set of all pure strategies that appear in some history in H'. H' is an *MCUBR configuration* if $S(H')$ is an MCUBR set. I can now state the main theorem of the paper.

Theorem 1 *Let Γ be a finite n-player game.*
i) If s/m is sufficiently small, the unperturbed process $P^{\gamma,m,s,0}$ converges with probability one to a minimal γ-configuration.
i) If, in addition, ε is sufficiently small, the perturbed process $P^{\gamma,m,s,\varepsilon}$ puts an arbitrarily high probability on the minimal γ-configuration(s) with minimum stochastic potential.

PROOF: See the Appendix.

Theorem 1 resembles Theorem 7.2 in Young (1998, p. 111), but does not require the game to be nondegenerate in any sense. This simplifies the first part of the proof which, in other respects, is analogous to Young's proof. Young requires that the game belong to the generic class of games with the following property.

Definition 4 *(Young, 1998) Γ is nondegenerate in best replies (NDBR) if for every i and every x_i, either $\beta_i^{-1}(x_i)$ is empty or it contains a nonempty subset that is open in the relative topology of $\square(X_{-i})$.*

2.4 Other Results

I will now proceed by demonstrating that two other results for β-adaptive play in Young (1998) can be obtained for γ-adaptive play in a straightforward manner. These results follow since γ-adaptive play is a *regular perturbed Markov process* (see the Appendix) and since, by the definition of the better-reply correspondence $\beta(p) \subset \gamma(p)$, and $\gamma(p) = x$ with $p(x) = 1$, if and only if x is a strict Nash equilibrium.

Young shows that a state h is an absorbing state of β-adaptive play, if and only if it consists of a strict Nash equilibrium played m times in succession. He calls such a state a *convention.* Due to the two last properties mentioned above, the same relation also holds for γ-adaptive play. It is clear that an absorbing state of this process must be a state on the form $h_x = (x, ..., x)$. Moreover, x must be a unique better reply to a distribution p with $p(x) = 1$ and hence, a strict Nash equilibrium. Conversely, any state consisting of m repetitions of a strict Nash equilibrium is clearly an absorbing state.

2.4.1 2×2 Coordination Games

Consider the game in Figure 2.2. This game is a 2×2 *Coordination Game* if (A, a) and (B, b) are strict Nash equilibria.

	a	b
A	$\pi_1(A,a), \pi_2(A,a)$	$\pi_1(A,b), \pi_2(A,b)$
B	$\pi_1(B,a), \pi_2(B,a)$	$\pi_1(B,b), \pi_2(B,b)$

FIGURE 2.2.

An equilibrium (A, a) of a 2×2 Coordination Game is *risk dominant* if its Nash product exceeds that of (B, b), or in other words if

$$\begin{aligned} & [\pi_1(A,a) - \pi_1(B,a)]\,[\pi_2(A,a) - \pi_2(A,b)] \\ \geq\ & [\pi_1(B,b) - \pi_1(A,b)]\,[\pi_2(B,b) - \pi_2(B,a)]\,. \end{aligned} \tag{2.3}$$

This definition (with a strict inequality) is due to Harsanyi and Selten (1988). I say that a convention $h = (x, ..., x)$ is *risk dominant,* if the pure-strategy profile x is a risk-dominant equilibrium.

Theorem 2 *Let Γ be a 2×2 Coordination Game.*
i) If information is sufficiently incomplete ($s/m \leq 1/2$) then, from any initial state, the unperturbed process $P^{\gamma,m,s,0}$ converges with probability one to a convention and locks in.
ii) If information is sufficiently incomplete ($s/m \leq 1/2$) and s sufficiently large, the stochastically stable states of the perturbed process $P^{\gamma,m,s,\varepsilon}$ correspond one to one with the risk-dominant conventions.

PROOF: Replace the words "best reply" and "best replies" by "better reply" and "better replies," and delete "therefore" on line 31 p. 69 in the proof of Theorem 4.2 in Young (1998, pp. 68-70). *Q.E.D.*

2.4.2 Weakly Acyclic Games

Given the n-player game Γ with finite strategy space $X = \prod X_i$, associate each pure-strategy profile $x \in X$ with the vertex of a graph. Draw a directed edge from vertex x to vertex x' if and only if both of the following conditions hold:

i) there exists exactly one player i such that $x_i \neq x_i'$ and

$$\pi_i\left(x_i', x_{-i}\right) \geq \pi_i\left(x\right),$$

ii) there does not exist $x_i'' \neq x_i$ and $x_i'' \neq x_i'$, such that

$$\pi_i\left(x_i', x_{-i}\right) > \pi_i\left(x_i'', x_{-i}\right) > \pi_i\left(x\right).$$

The graph obtained in this manner is called a *better-reply graph* (Josephson and Matros, 2000). A *better-reply path* is a sequence of the form $x^1, x^2, ..., x^l$ such that each pair (x^j, x^{j+1}) corresponds to a directed edge in the better-reply graph. I say that a game is *weakly acyclic in* γ if there exists a directed path to a *sink*, a vertex with no outgoing edges, from every vertex.

Theorem 3 *Let Γ be a finite n-player game that is weakly acyclic in γ.*
i) If s/m is sufficiently small, the unperturbed process $P^{\gamma,m,s,0}$ converges with probability one to a convention from any initial state.
ii) If, in addition, ε is sufficiently small, the perturbed process puts arbitrarily high probability on the convention(s) with minimum stochastic potential.

PROOF: Replace $P^{m,s,o}$ by $P^{\gamma,m,s,o}$, and the words "best replies" and "best-reply path" by "better replies" and "better-reply path," respectively in the proof of Theorem 7.1 in Young (1998, pp. 163-164). *Q.E.D.*

Clearly, if a game is weakly acyclic in β (that is, weakly acyclic according to Young's (1998) definition), then it is also weakly acyclic in γ. The opposite may not hold, as illustrated by the game in Figure 2.3. In this game, there is a better-reply path from any vertex to $\{C\} \times \{c\}$, but no best-reply path from any vertex in $\{A, B\} \times \{a, b\}$ to this unique sink.

	a	b	c
A	2, 1	1, 2	0, 0
B	0, 2	2, 1	0, 0
C	1, 0	0, 0	2, 2

FIGURE 2.3

2.5 Relation to Regular Selection Dynamics

In this section, I relate the above result concerning the connection between minimal sets closed under better replies and stochastically stable states under γ-adaptive play to Ritzberger and Weibull's (1995) findings in a deterministic continuous-time model. Apart from the stochastic element in my model, there is another important difference between the approaches in these two models. Whereas I make detailed assumptions about individual behavior in the sense that all individuals are assumed to apply a specific adaptive rule, Ritzberger and Weibull (1995) make general assumptions about aggregate population dynamics. Yet, there is a clear connection between the results in the two models, as will be discussed below. I first reprint two definitions and an important theorem from Ritzberger and Weibull (1995).

Definition 5 *(Ritzberger and Weibull, 1995) A* ***regular selection dynamics*** *on* $\square(X)$ *is a system of ordinary differential equations:* $\dot{p}_i^k = f_i^k(p)p_i^k$ $\forall k = 1, ..., |X_i|$, $\forall i \in N$ *with* $f_i : \square(X) \to \mathbb{R}^{|X_i|}$ *and* $f = \prod_{i \in N} f_i$ *is such that*
(i) f *is Lipschitz continuous on* $\square(X)$
(ii) $f_i(p) \cdot p_i = 0$ $\forall p \in \square(X)$, $\forall i \in N$.

Definition 6 *(Ritzberger and Weibull, 1995) A* ***sign-preserving selection dynamics (SPS)*** *is a regular selection dynamics such that for all* $i \in N$, *all* $p \in \square(X)$, *and all* x_i^k, *such that* $p_i(x_i^k) > 0$,

$$u_i(x_i^k, p_{-i}) < u_i(p) \Leftrightarrow f_i^k(p) < 0.$$

Theorem 4 *(Ritzberger and Weibull, 1995) For any SPS dynamics and any set* $Y \in \mathcal{X}$, $\square(Y)$ *is asymptotically stable if and only if* Y *is closed under better replies.*

I can make two general observations by comparing Theorem 1 with Theorem 4. First, my result is consistent with that for SPS dynamics in the sense that the limiting distribution of γ-adaptive play puts positive probability only on histories with support in sets which are asymptotically stable in SPS dynamics. Second, my result provides a tool for selection among different asymptotically stable sets. The unperturbed process selects those asymptotically stable sets which correspond to minimal closed sets under better replies and the perturbed process selects those minimal closed set(s) that minimize the stochastic potential.

In interpreting the above results, it is important to keep in mind the distinction between asymptotic and stochastic stability. Asymptotic stability refers to robustness against small one-time shocks. It is a local property since it only assures that nearby points will converge to the stable set. Stochastic stability refers to robustness against perpetual random shocks. It is a global property in the sense that the perturbed process, independently of the initial state, will spend most of the time in the stochastically stable states in the long run.

2.6 Examples

Consider the two-player game in Figure 2.4, taken from Young (1998). This game has two closed sets under better replies, X and $\{D\} \times \{d\}$, of which the latter is a minimal set. This set is also the unique minimal closed set under best replies. Young shows that this game is nondegenerate in best replies, since strategy B is a best reply to a set of distributions which is neither empty nor contains a set open in $\square(X_2)$. The unperturbed version of β-adaptive play has two recurrent classes, one with span $\{A, C\} \times \{a, c\}$ and one with span $\{D\} \times \{d\}$. However, it is easy to see that B is a better reply to several pure-strategy profiles and, by Theorem 1 γ-adaptive play converges with probability one to the state $h = ((D, d), ..., (D, d))$. By Theorem 4, $\{D\} \times \{d\}$ is asymptotically stable under SPS dynamics.

	a	b	c	d
A	$0,1$	$0,0$	$2,0$	$0,0$
B	$2/(1+\sqrt{2}),0$	$-1,1/2$	$2/(1+\sqrt{2}),0$	$0,0$
C	$2,0$	$0,0$	$0,1$	$0,0$
D	$0,0$	$1,0$	$0,0$	$2,2$

FIGURE 2.4

The game in Figure 2.5 is nondegenerate in β. It has four sets closed under β: X, $Y = \{B\} \times \{b\}$ and $Z = \{A, B\} \times \{a, b\}$ and $T = \{C\} \times \{c\}$. It has three sets closed under γ: X, Y and T. The two minimal closed sets under β are identical to the two minimal closed set under γ, Y and T. According to Theorem 4, $\square(X)$, $\square(Y)$ and $\square(T)$ are asymptotically stable under SPS dynamics. By Theorem 1, it follows that the unperturbed version of γ-adaptive play converges with probability one to either Y or T. Moreover, it is easy to check that T has minimum stochastic potential, implying that the state $h = ((C, c), ..., (C, c))$ is stochastically stable.

	a	b	c
A	2, 5	2, 5	0, 0
B	0, 0	3, 1	0, 0
C	1, 3	0, 0	4, 4

FIGURE 2.5

A minimal closed set under γ always contains a minimal closed set under β, but as the following example illustrates, the sets of stochastically stable states under the two associated dynamics do not necessarily intersect. The game in Figure 2.6 has two minimal sets closed under best replies, $\{A, B\} \times \{a, b\}$ and $\{C\} \times \{c\}$, and one set closed under better replies, $\{C\} \times \{c\}$. However, whereas only pure-strategy profiles in $\{A, B\} \times \{a, b\}$ are played with positive probability under the limiting distribution of β-adaptive play, the unique stochastically stable state under γ-adaptive play is a repetition of $\{C\} \times \{c\}$.

	a	b	c
A	1, 10	10, 1	0, 0
B	10, 1	1, 10	0, 0
C	0, 0	2, 0	1, 1

FIGURE 2.6

2.7 Extensions

The better-reply and the best-reply correspondences both belong to the class of *behavior correspondences* (Ritzberger and Weibull, 1995). Behavior correspondences are upper hemi-continuous correspondences, such that the image of any mixed-strategy profile includes the set of best-replies. I believe that the main result in this paper can be extended in a straight-forward manner to other correspondences in this class if the game is assumed to be non-degenerate in a specific sense. In other words, my conjecture is that each behavior correspondence φ gives rise to a regular perturbed Markov process and that the unperturbed version of this process converges with probability one to a minimal φ-configuration. If, in addition, the mutation rate is sufficiently small, the perturbed version of the process puts arbitrarily high probability on the minimal φ-configuration(s) with minimum stochastic potential. This result requires that the set of replies to all sample distributions over a particular set be identical to the set of replies to all possible distributions over the same set. For some correspondences, such as the best-reply correspondence, this is assured if it is assumed that the game fulfils certain conditions. For other correspondences, like the better-reply correspondence, such an assumption is not needed.

2.8 Conclusion

In this paper I analyze a model where individuals are recurrently matched to play a finite n-player game and almost always play a myopic better reply to a sample distribution from the recent history of play. I prove that in all finite games, as the mutation rate tends to zero, only pure strategies in certain minimal closed sets under better replies will be played with positive probability. This result extends Young's (1993, 1998) result for myopic best replies in generic games. It is also consistent with Ritzberger and Weibull's (1993) result on the equivalence of asymptotically stable sets and closed sets under better replies, for a wide class of deterministic continuous-time dynamics. Finally, I show that the risk-dominant equilibria are selected in 2×2 Coordination Games, and certain strict Nash equilibria in weakly acyclic games.

2.9 Appendix

Definition 7 *(Young, 1998) P^ε is a* **regular perturbed Markov process** *if P^ε is irreducible for every $\varepsilon \in (0, \varepsilon^*]$, and for every $h, h' \in H$, $P^\varepsilon_{hh'}$ approaches $P^0_{hh'}$ at an exponential rate, i.e. $\lim_{\varepsilon\to 0} P^\varepsilon_{hh'} = P^0_{hh'}$ and if $P^\varepsilon_{hh'} > 0$ for some $\varepsilon > 0$, then $0 < \lim_{\varepsilon\to 0} \frac{P^\varepsilon_{hh'}}{\varepsilon^{r_{hh'}}} < \infty$ for some $r_{hh'} \geq 0$.*

Lemma 2 *γ-adaptive play is a regular perturbed Markov process.*

PROOF: This proof follows Young (1998, p. 55) closely. The process $P^{\gamma,m,s,\varepsilon}$ operates on the finite space X^m of length-m histories. Given a history $h = (x^1, x^2, ..., x^m)$ at time t, the process moves to a state of form $h' = (x^2, x^3..., x^m, x)$ for some $x \in X$ in the next period. Recall that any such state h' is said to be a successor of h. Before choosing a pure strategy, an individual in role i draws a sample of size s from the m previous choices in h for each role $j \in \{1, ..., n\}$ (including her own), the samples being independent among distinct player roles j. The action x_i is an idiosyncratic choice or error if and only if there exists no set of n samples in h (one for each player role j), such that $x_i \in \gamma_i(\hat{p})$, where $\hat{p}$ is the product of the frequency distributions. For each successor h', let $r_{hh'}$ denote the total number of errors in the rightmost element of h'. Evidently, $0 < r_{hh'} < n$. It is easy to see that the probability of the transition $h \to h'$ is in the order of $\varepsilon^{r_{hh'}}(1-\varepsilon)^{n-r_{hh'}}$. If h' is not a successor of h, the probability of the transition $h \to h'$ is zero. Thus, the process $P^{\gamma,m,s,\varepsilon}$ approaches $P^{\gamma,m,s,0}$ at a rate approximately exponential in ε; furthermore it is irreducible whenever $\varepsilon > 0$. It follows that $P^{\gamma,m,s,\varepsilon}$ is a regular perturbed Markov process. *Q.E.D.*

Theorem 5 *(Young, 1998) Let P^ε be a regular perturbed Markov process and let μ^ε be the unique stationary distribution of P^ε for $\varepsilon > 0$. Then $\lim_{\varepsilon\to 0} \mu^\varepsilon = \mu^0$ exists, and is a stationary distribution of P^0. The stochastically stable states are precisely those states that are contained in the recurrent classes of P^0 having minimum stochastic potential.*

PROOF OF THEOREM 1: Except for the first part, this proof is analogous to the proof of Theorem 7.2 in Young (1998, pp. 164-166). In order to prove the first claim, I show that for a sufficiently large s and sufficiently small s/m, the spans of the recurrent classes of the unperturbed process correspond one to one with the minimal sets closed under γ of the game.

First, note that by Corollary 1 $\gamma_i(\Box(Y)) = \tilde{\gamma}_i(Y)$ for all $Y \in \mathcal{X}$. This implies that the set of better replies to the distributions on an arbitrary set of pure-strategy profiles is identical to the set of better replies to the sample distributions on the same set.

For each product set Y and player i, define the mapping $\check{\gamma}_i(Y) = Y_i \cup \gamma_i(\Box(Y))$ and $\check{\gamma}(Y) = \prod \check{\gamma}_i(Y)$. Choose m such that $m \geq s\,|X|$. Furthermore, fix a recurrent class E_k of $P^{\gamma,m,s,0}$, and choose any $h^0 \in E_k$ as the initial state. I shall show that the span of E_k, $S(E_k)$, is a minimal closed set under γ. It is evident that there is a positive probability of reaching a state h^1 in which the most recent s entries involve a repetition of some fixed $x^* \in X$. Note that $h^1 \in E_k$, because E_k is a recurrent class. Let $\check{\gamma}^{(j)}$ denote the j-fold iteration of $\check{\gamma}$ and consider the nested sequence:

$$\{x^*\} \subseteq \check{\gamma}(\{x^*\}) \subseteq \check{\gamma}^{(2)}(\{x^*\}) \subseteq ... \subseteq \check{\gamma}^{(j)}(\{x^*\}) \subseteq ... \tag{2.4}$$

Since X is finite, there exists some point at which this sequence becomes constant, say, $\check{\gamma}^{(j)}(\{x^*\}) = \check{\gamma}^{(j+1)}(\{x^*\}) = Y^*$. By construction, Y^* is a set closed under γ.

Assume $\check{\gamma}(\{x^*\}) \neq \{x^*\}$ (otherwise the following argument is redundant). Then, there is a positive probability that, beginning after the history h^1, some $x^1 \in \check{\gamma}(\{x^*\})\backslash\{x^*\}$ will be chosen for the next s periods. Call the resulting history h^2. Then, there is a positive probability that $x^2 \in \check{\gamma}(\{x^*\})\backslash\{x^*, x^1\}$ will be chosen for the next s periods and so forth. Continuing in this way, one eventually obtains a history h^k such that all elements of $\check{\gamma}(\{x^*\})$, including the original $\{x^*\}$, appear at least s times. All that is needed to assume is that m is large enough, so that the original s repetitions of x^* have not been forgotten. This is assured if $m \geq s\,|X|$. Continuing this argument, it is clear that there is a positive probability of eventually obtaining a history h^*, where all members of Y^* appear at least s times within the last $s\,|Y|$ periods. In particular, $S(h^*)$ contains Y^* which, by construction, is a set closed under γ.

I claim that Y^* is a minimal closed set under γ. Let Z^* be a minimal closed set under γ contained in Y^*, and choose $z^* \in Z^*$. Beginning with the history h^* already constructed, there is a positive probability that z^* will be chosen for the next s periods. After this, there is a positive probability that only elements of $\check{\gamma}(\{z^*\})$ will be chosen, or members of $\check{\gamma}^2(\{x^*\})$, or members of $\check{\gamma}^3(\{x^*\})$, and so on. This happens if individuals always draw samples from the new part of the history that follows h^*, which they will do with positive probability.

The sequence $\check{\gamma}^{(k)}(\{z^*\})$ eventually becomes constant with value Z^*, because Z^* is a minimal closed set under γ. Moreover, the part of the history before the s-fold repetition of x^* will be forgotten within m periods. Thus, there is a positive probability of obtaining a history h^{**} such that $S(h^{**}) \subseteq Z^*$. From such a history, the process $P^{m,s,0}$ can never generate a history with members that are not in Z^* because Z^* is a set closed under γ.

Since the chain of events that led to h^{**} began with a state in E_k, which is a recurrent class, h^{**} is also in E_k; moreover, every state in E_k is reachable from h^{**}. It follows that $Y^* \subseteq S(E_k) \subseteq Z^*$, from which it can be concluded that $Y^* = S(E_k) = Z^*$.

Conversely, I must show that if Y^* is a minimal closed set under γ, then $Y^* = S(E_k)$ for some recurrent class E_k of $P^{\gamma,m,s,0}$. Choose an initial history h^0 that involves only pure-strategy profiles in Y^*. Starting at h^0, the process $P^{\gamma,m,s,0}$ generates histories involving no pure-strategy profiles outside of $S(h^0)$, $\check{\gamma}(S(h^0))$, $\check{\gamma}^{(2)}(S(h^0))$ and so on. Since Y^* is a set closed under γ, all these pure-strategy profiles must occur in Y^*. With probability one, the process eventually enters a recurrent class, say E_k. It follows that $S(E_k) \subseteq Y^*$. Since Y^* is a minimal closed set under γ, the earlier part of the argument shows that $S(E_k) = Y^*$. This establishes the one to one correspondence between minimal sets closed under γ and the recurrent classes of $P^{\gamma,m,s,0}$.

The second claim of Theorem 1 now follows from the fact that γ-adaptive play, by Lemma 2, is a regular perturbed Markov process, and by Theorem 5, which states that the stochastically stable states of such a process are the states contained in the recurrent classes of the unperturbed process with the minimum stochastic potential. *Q.E.D.*

Bibliography

[1] Foster, D., and P. Young (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* 38: 219-32.

[2] Freidlin, M., and A. Wentzell (1984): *Random Perturbations of Dynamical Systems.* Springer-Verlag.

[3] Harsanyi, J., and R. Selten (1988): *A General Theory of Equilibrium Selection in Games.* The MIT Press.

[4] Josephson, J., and A. Matros (2000): "Stochastic Imitation in Finite Games," SSE/EFI Working Paper Series in Economics and Finance No. 363, Stockholm.

[5] Ritzberger, K., and J. W. Weibull (1995): "Evolutionary Selection in Normal-Form Games," *Econometrica* 63, 1371-1400.

[6] Simon, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69, 99-118.

[7] Young, P. H. (1993): "The Evolution of Conventions," *Econometrica* 61, 57-84.

[8] —— (1998): *Individual Strategy and Social Structure.* Princeton University Press.

Chapter 3

Stochastic Adaptation in Finite Games Played by Heterogeneous Populations

3.1 Introduction

There is an extensive literature on learning in games, which investigates the long-run outcomes when boundedly rational individuals use simple rules of adaptation, or *learning rules*, to play games.[1] A common assumption in this literature is that all individuals employ the same learning rule. In this paper, I depart from this, in many cases unrealistic, assumption and allow for several different learning rules in each population.

More specifically, I analyze stochastic learning in finite n-player games with heterogeneous populations of myopic *best repliers*, *better repliers* and *imitators*. The best repliers observe a sample from a finite history of their opponents' past play, calculate an empirical distribution, and choose a best reply to this distribution. The better repliers observe a sample from a finite history of their opponents' and their own population's past play, and choose any pure strategy which gives at least as high expected payoff against the empirical distribution of the opponents' play. This behavior can be seen as a special case of Simon's (1955) satisfying behavior. Imitators, finally, observe a sample of their own population's past play and payoffs and either choose a pure strategy with maximum average payoff, or a pure strategy with maximum payoff.[2] An alternative interpretation of this framework is that each

[1] See Fudenberg and Levine (1998) for an introduction.

[2] For an analysis of stochastic adaptation by homogeneous populations of better repliers, see Josephson (2000), and by homogeneous populations of imitators, see Josephson and

individual chooses a mixed strategy with support on the set of pure strategies that are either best replies, better replies, or consistent with imitation. For experimental support of such an intermediate rule, see Camerer and Ho (1999) and Stahl (2000).

I prove that for any finite game, if the ratio between the sample and history size is sufficiently small, the resulting unperturbed Markov chain converges with probability one to a *minimal closed set under better replies* of the game. This result is independent of population shares, as long as the share of better repliers is positive, and implies convergence to a strict Nash equilibrium in weakly acyclical games. It is also consistent with previous results in a different framework. Ritzberger and Weibull (1995) show that for a large class of deterministic selection dynamics in continuous time, including the replicator dynamics, a product set is asymptotically stable if and only if it is a minimal closed set under better replies.

I then perturb the stochastic process by assuming that with a small probability, the individuals make mistakes or experiment and play a pure strategy at random according to some fixed probability distribution with full support. A mathematically equivalent interpretation is that each population contains a small share of noise players who always choose a pure strategy at random. This assumption makes the process irreducible and aperiodic, and thus implies a unique stationary distribution. I calculate the support of this distribution as the probability of mistakes tends to zero. If the sample size, and the information incompleteness is sufficiently large, and if one minimal closed set under better replies strictly Pareto dominates all strategy-tuples outside the set, then the perturbed Markov chain puts probability one on this set as the level of noise tends to zero. A corollary of this is that the strictly Pareto-dominant equilibrium is selected in games of common interest and 2×2 Coordination Games. However, the minimum sample size required for this result depends on the payoffs of the game. In fact, for sample sizes below a certain critical level, the *risk-dominant* equilibrium is always selected in all symmetric Coordination Games, and also in many asymmetric Coordination Games, a finding which seems to be consistent with experimental evidence (see Van Huyck, 1997).

The basic setting in this paper is similar to that in Young (1993a,1998), who analyzes the dynamics of a homogenous population of myopic best repliers. The present paper is also related to earlier papers on stochastic learning with heterogeneous populations, but it differs in that the analysis is applied to general finite games, and not only special classes of such games. Kaniovski, Kryazhimiskii, and Young (2000) study adaptive dynamics in 2×2

Matros (2000).

Coordination Games played by heterogeneous populations of myopic best repliers, conformists (who do what the majority does), and nonconformists (who do the opposite of what the majority does). They show that the resulting process may have limit cycles even when the proportion of non-best repliers is arbitrary small. Kaarbøe and Tieman (1999) use the Kandori, Mailath, and Rob (1993) framework to study strictly supermodular games played by best repliers and imitators. In their model, the Pareto efficient equilibrium is selected. In Weibull and Sáez Marti (1998), Young's (1993b) evolutionary version of Nash's demand game is played by one population of myopic best repliers only, and another population of best repliers and clever agents who play a best reply to the best reply. The authors show that Young's predictions are still valid in the presence of such clever agents. Matros (2000) extends this result to finite two-player games.

This paper is organized as follows. I start by describing the model in Section 2. In Section 3, I present the general results and in Section 4, the results for 2×2 Coordination Games. Section 5 concludes. Omitted proofs can be found in the Appendix.

3.2 Model

The basic setting is similar to that of Young (1993a, 1998), although my notation is slightly different. Let Γ be a finite n-player game in strategic form. Let X_i be the finite set of pure strategies x_i available to player $i \in \{1, ..., n\} = N$ and let $\Delta(X_i)$ be the set of probability distributions p_i over these pure strategies. Define the product sets $X = \prod X_i$ and $\Box(X) = \prod_i \Delta(X_i)$ with typical elements x and p, respectively. Let $p_i(x_i)$ denote the probability mass on pure strategy x_i and let $p(x) = \prod_{i \in N} p_i(x_i)$. I write $x_{-i} \in \prod_{j \neq i} X_j = X_{-i}$ and $p_{-i} \in \prod_{j \neq i} \Delta(X_j) = \Box(X_{-i})$ to represent the pure strategies and distributions of the pure strategies of player i's opponents.

Let $C_1, ..., C_n$ be n finite and non-empty populations, each consisting of a share $\delta_i^\alpha > 0$ of *imitators*, a share $\delta_i^\beta > 0$ of *best repliers*, and a share $\delta_i^\gamma > 0$ of *better repliers* (all to be defined below). The populations need not be of equal size, nor do they necessarily have equal shares of learning rules. Each member of population C_i is a candidate to play role i in the game G and has payoffs represented by the utility function $\pi_i : X \to \mathbb{R}$, and expected payoffs represented by the function $u_i : \Box(X) \to \mathbb{R}$. In slight abuse of notation, I write $u_i(x_i, p_{-i})$ instead of $u_i(p_i, p_{-i})$ if $p_i(x_i) = 1$.

Let $t = 1, 2, ...$ denote successive time periods. The stage game Γ is played once in each period. In period t, one individual is drawn at random from each of the n populations and assigned to play the corresponding role.

The individual in role i chooses a pure strategy x_i^t from her strategy space X_i, according to her personal decision rule (to be defined below). The pure-strategy profile $x^t = (x_1^t, ..., x_n^t)$ is recorded and referred to as *play* at time t. The *history* or *state* at time $t \geq m$ is the sequence $h^t = (x^{t-m+1}, ..., x^t)$, where m is the *memory size* of all individuals. Let $H = X^m$ be the finite set histories of length m and let h be an arbitrary element of this set. Assume an arbitrary initial history $h^m = (x^1, ..., x^m)$ at time m.

Strategies are chosen as follows. In every period $t > m$, each individual drawn to play the game inspects a sample of size s drawn without replacement from the most recent m periods.[3] The draws are independent for the various individuals and across time. If an imitator is drawn, she observes a sample of size s of population-specific strategy and payoff realizations and chooses a pure strategy which is included in the sample, and has maximum average payoff or, alternatively, maximum empirical payoff. If a best replier is drawn, she inspects a sample of size s, consisting of $n-1$ independent draws of her opponents' strategy realizations and calculates an empirical distribution, $\hat{p}_{-i}$. she then chooses a pure strategy which is a best reply to this distribution. If a better replier is drawn, she inspects a sample of size s, consisting of n independent draws, of all players' strategy realizations and calculates an empirical distribution, $\hat{p}$. Thereafter, she chooses a pure strategy which gives at least as high expected payoff against the empirical distribution of the opponents' strategies $\hat{p}_{-i}$ as $\hat{p}_i$. Formally, she chooses a pure strategy $x_i \in \gamma_i(\hat{p})$, where γ_i is the better-reply correspondence, defined by

$$\gamma_i(p) = \{x_i \in X_i \mid u_i(x_i, p_{-i}) \geq u_i(p)\}. \tag{3.1}$$

For each $x_i \in X_i$, let $p_i(x_i \mid h)$ be the conditional probability that the individual in role i chooses pure strategy x_i, given history h. I assume that $p_i(x_i \mid h)$ is independent of t, and that $p_i(x_i \mid h) > 0$ if and only if at least one of the following conditions hold:

i) There exists a sample of size s from player i's m most recent strategy and payoff realizations, which includes x_i, and where x_i has the maximum average realized payoff.

ii) There exists a sample of size s from player i's m most recent strategy and payoff realizations, which includes x_i, and where x_i has the maximum realized payoff.

iii) There exists a sample of size s from history h, consisting of $n-1$ independent draws and with a sample distribution of $\hat{p}_{-i} \in \square(X_{-i})$, such that $x_i \in \beta_i(\hat{p}_{-i})$, where β_i is the best-reply correspondence.

[3] In fact, the sample size may differ among individuals, as long as the *minimum* sample size in each share of the population is equal to s.

iv) There exists a sample of size s from the history h, consisting of n independent draws and with a sample distribution of $\hat{p} \in \square(X)$, such that $x_i \in \gamma_i(\hat{p})$, where γ_i is the better-reply correspondence defined above.

Given a history $h^t = (x^{t-m+1}, ..., x^t)$ at time t, the process moves to a state of the form $h^{t+1} = (x^{t-m+2}, ..., x^t, x^{t+1})$ in the next period. Such a state is called a *successor* of h^t. My behavioral assumptions imply that the process moves from a current state h to a successor state h' in each period, according to the following transition rule. If x is the rightmost element of h, the probability of moving from h to h' is $P_{hh'}^{*,m,s,0} = \prod_{i=1}^{n} p_i(x_i \mid h)$ if h' is a successor of h and 0 if h' is not a successor of h. This defines a finite Markov chain on the finite state space of histories H. I denote the process $P^{*,m,s,0}$ *heterogeneous adaptive play with memory m and sample size s*, but often refer to it as the *unperturbed process.*

Several concepts from Markov chain theory are used below. A *recurrent class* E_k of the process $P^{*,m,s,0}$ is a set of states such that there is zero probability of moving from any state in the class to any state outside, and a positive probability of moving from any state in the class to any other state in the class in a finite number of periods. A state h is *absorbing* if it constitutes a singleton recurrent class. The *basin of attraction* of a state h' is the set of states h such that there is a positive probability of moving from h to h' in a finite number of periods under the unperturbed process $P^{*,m,s,0}$.

I also define a perturbed process in a manner completely analogous to Young (1993a, 1998). Formally, in each period there is some small probability ε that the individual in role i experiments or makes a mistake and chooses a pure strategy at random from X_i, instead of according to her decision rule. The event that the individual in role i experiments is assumed to be independent of the event that the individual in role j experiments for every $j \neq i$. For every role i, let $q_i(x_i \mid h)$ be the conditional probability that the individual in role i chooses $x_i \in X_i$, given that she experiments and the process is in state h. I assume that $q_i(x_i \mid h)$ is independent of t and decision rules, and that $q_i(x_i \mid h) > 0$ for all $x_i \in X_i$. Suppose that the process is in state h at time t. Let J be a subset of j players. The probability is $\varepsilon^j(1-\varepsilon)^{n-j}$ that exactly the players in J experiment and the others do not. Conditional on this event, the transition probability of moving from h to h' is $Q_{hh'}^J = \prod_{i \in J} q_i(x_i \mid h) \prod_{i \notin J} p_i(x_i \mid h)$ if h' is a successor of h and x is the rightmost element of h' and 0 if h' is not a successor of h. This gives the following transition probability of the perturbed Markov process:

$$P_{hh'}^{*,m,s,\varepsilon} = (1-\varepsilon)^n P_{hh'}^{*,m,s,0} + \sum_{J \subset N, J \neq \emptyset} \varepsilon^{|J|}(1-\varepsilon)^{n-|J|} Q_{hh'}^J.$$

I call the process $P^{*,m,s,\varepsilon}$ *heterogeneous adaptive play with memory m, sample* *size s, experimentation probability ε and experimentation distributions q_i.*

This process is irreducible and aperiodic and thus, has a unique stationary distribution μ^ε. I study this distribution as ε tends to zero. In my analysis, I use the following standard definitions, due to Freidlin and Wentzell (1984) and Foster and Young (1990). A state h is *stochastically stable* if $\lim_{\varepsilon \downarrow 0} \mu^\varepsilon(h)$ exists and is positive. For each pair of distinct recurrent classes, E_k and E_l, a *kl-path* is a sequence of states $\zeta = (h^1, h^2..., h^q)$ beginning in E_k and ending in E_l. The *resistance* of this path is the sum of the resistances on the edges that compose it. Let r_{kl} be the least resistance over all kl-paths. Construct a complete directed graph with one vertex for each recurrent class. The weight on the directed edge $E_k \rightarrow E_l$ is r_{kl}. A *tree* rooted at E_l is a set of directed edges such that, from every vertex different from E_l, there is a unique directed path in the tree to E_l. The *resistance* of a rooted tree is the sum of the resistances on the edges that compose it. The *stochastic potential* $\rho(E_l)$ of a recurrent class E_l is the minimum resistance over all trees rooted at E_l.

3.3 General Results

In this section, I will present two general results on the asymptotic distribution of heterogeneous adaptive play: one theorem for all finite n-player games, and one theorem for finite n-player games that can be Pareto ranked in a special sense. In order to state these theorems, some further definitions are needed. Let $\mathcal{X}$ be the collection of all non-empty product sets $Y \subset X$. Let $\Delta(Y_i)$ be the set of probability distribution with support in Y_i and let $\square(Y) = \prod_{i \in N} \Delta(Y_i)$ be the corresponding product set.

Definition 1 *(Ritzberger and Weibull, 1995) A set $Y \in X$ is* ***closed under better replies (CUBR)*** *if $\gamma(\square(Y)) \subseteq Y$. A set $Y \in X$ is a* ***minimal closed set under better replies (MCUBR)*** *if it is closed under better replies and contains no proper subset with this property.*

Let $H' \subseteq H$ be an arbitrary set of histories. Define the *span* of H', $S(H')$, as the product set of all pure strategies appearing in some history of H' and let H^Y be the set of all histories h such that $S(\{h\}) \subseteq Y$. I say that a set of histories H' is *forward invariant* if the Markov chain remains forever in H', once it has reached a state in this set.

Lemma 1 *For any MCUBR set Y, H^Y is forward invariant under $P^{*,m,s,0}$.*

PROOF: It is evident that imitation of the best pure strategy in a sample containing only pure strategies in an MCUBR set cannot result in the play of a pure strategy outside the MCUBR set. Moreover, by definition, a best or better reply to a probability distribution on an MCUBR set is always contained in the MCUBR set. Hence, if the unperturbed Markov chain $P^{*,m,s,0}$ is in a state consisting only of pure-strategy profiles in a particular MCUBR set, then a pure strategy outside the MCUBR set will thereafter never be played by any of the individuals drawn to play. *Q.E.D.*

In order to state the main theorem, one more definition is needed. I say that a set of states H' is a *minimal γ-configuration* if $S(H')$ is an MCUBR set. In other words, H' is a minimal γ-configuration if all the strategies in some MCUBR set Y, and no other strategies, appear in the histories of H'.

Theorem 1 *Let Γ be a finite n-player game.*
i) If s/m is sufficiently small, the unperturbed process $P^{*,m,s,0}$ converges with probability one to a minimal γ-configuration.
i) If, in addition, ε is sufficiently small, the perturbed process $P^{*,m,s,\varepsilon}$ puts an arbitrarily high probability on the minimal γ-configuration(s) with minimum stochastic potential.

PROOF: See the Appendix.

This theorem is analogous to Theorem 1 in Josephson (2000), which states that in a homogeneous setting, where the share of better repliers in each population is one, for a sufficiently large sample size and sufficiently incomplete information ($s/m \leq \frac{1}{|X|}$), the unperturbed process converges with probability one to a minimal γ-configuration. The proof of Theorem 1 is based on this result combined with Lemma 1, and the observation that in each period, there is a positive probability that only better repliers will be drawn to play from the heterogeneous populations.

Theorem 1 is also consistent with previous results for deterministic continuous time selection dynamics. Ritzberger and Weibull (1995) show that a set is MCURB if and only if it is asymptotically stable for regular selection dynamics that are *sign preserving*. This is a large class of selection dynamics, which contain several well know dynamics, such as the replicator dynamics.

It is worth noting that Theorem 1 makes no reference to population mixtures. As long as the share of better repliers is positive in all populations, the span of the recurrent sets of the unperturbed process is independent of the shares of different learning rules in the populations. However, it is clear that this does not imply that the exact shape of the asymptotic distribution is independent of the population mixtures.

One may also note that in the class of games where all minimal closed sets under best replies and MCUBR sets coincide, it is sufficient if one of the population shares of better and best repliers is positive in each of the populations. This class of games contains, for instance, 2×2 Coordination Games.

In the special class of games where all MCUBR sets are singleton, and hence also strict Nash equilibria, Theorem 1 implies convergence to a convention, a monomorphic state which is a repetition of a strict Nash equilibrium. *Weakly acyclical* games (Young, 1993a, 1998), and games that are *weakly acyclical in* γ (Josephson, 2000) belong to this class.

The second result of this paper characterizes the stochastically stable states of heterogeneous play in finite n-player games with a Pareto optimal outcome.

Definition 2 *A non-empty set of strategy-tuples $M \subset X$* ***strictly Pareto dominates*** *a pure-strategy profile $y \in X \backslash M$ if, for all $i \in N$,*

$$\min_{x \in M} \pi_i(x) > \pi_i(y). \tag{3.2}$$

Theorem 2 *Let Γ be a finite n-player game with an MCUBR set Y, which strictly Pareto dominates all other pure-strategy profiles. If $s/m \leq 1/|X|$, s sufficiently large, and ε is sufficiently small, the perturbed process $P^{*,m,s,\varepsilon}$ puts arbitrarily high probability on a minimal γ-configuration with span Y.*

PROOF: See the Appendix.

The proof of Theorem 2 uses the following two observations. First, a state in a strictly Pareto-dominant set Y can be reached from any state outside the set if all individuals drawn to play simultaneously make a mistake and play a pure strategy in Y, and a sequence of imitators thereafter pick samples including this mistake. Second, the resistance of the reverse transition can be made arbitrarily large by choosing a sufficiently large sample and memory size.

Theorem 2 does not hold under the weaker condition that an MCUBR set strictly Pareto dominates all other MCUBR sets. Consider the game in Figure 3.1. In this game, there are two MCUBR sets, (A, a) and (C, c), and (A, a) strictly Pareto dominates (C, c). However, (A, a) does not strictly Pareto dominate (B, b), and for a sufficiently large sample size and sufficiently incomplete information, the stochastic potential is two for both $h_{(A,a)}$ and $h_{(C,c)}$. This follows since the process will make the transition from $h_{(A,a)}$ to $h_{(B,b)}$, if the two players simultaneously make mistakes in period t and play

(B, b), and a sequence of $s-1$ imitators in population C_1, corresponding to the row player, thereafter are drawn to play and sample only x^t and earlier strategy realizations, and finally imitators in both populations sample only from plays more recent than x^{t-1} for $m-s$ periods. Since state $h_{(B,b)}$ is clearly in the basin of attraction of $h_{(C,c)}$, this implies that the stochastic potential of $h_{(C,c)}$ is two at most.

	a	b	c
A	3, 3	0, 0	0, 0
B	0, 0	4, 0	1, 1
C	0, 0	0, 0	2, 2

FIGURE 3.1.

In *games of common interest* there exists a strict Nash equilibrium – a singleton MCUBR set – which strictly Pareto dominates all other strategy-tuples. From Theorem 2, it immediately follows that for a sufficiently large memory and sample size, and sufficiently incomplete information, this equilibrium is played with probability one under the limiting distribution.

3.4 Results for 2×2 Coordination Games

In this section, I will study the predictions of heterogeneous adaptive play in the special class of 2×2 Coordination Games and illustrate how the stochastically stable states depend on the sample size.

Consider the game in Figure 3.1.

	a	b
A	$\pi_1(A,a), \pi_2(A,a)$	$\pi_1(A,b), \pi_2(A,b)$
B	$\pi_1(B,a), \pi_2(B,a)$	$\pi_1(B,b), \pi_2(B,b)$

FIGURE 3.2.

This game is a 2×2 *Coordination Game* if (A, a) and (B, b) are strict Nash equilibria. It is a *symmetric* 2×2 *Coordination Game* if, in addition, the diagonal payoffs are equal for the two players, $\pi_1(B,a) = \pi_2(A,b)$ and $\pi_1(A,b) = \pi_2(B,a)$. An equilibrium (A,a) of a 2×2 Coordination Game is *risk dominant* if its Nash product exceeds that of (B,b):

$$\begin{aligned} & [\pi_1(A,a) - \pi_1(B,a)]\,[\pi_2(A,a) - \pi_2(A,b)] \\ \geq\ & [\pi_1(B,b) - \pi_1(A,b)]\,[\pi_2(B,b) - \pi_2(B,a)]\,. \end{aligned} \tag{3.3}$$

This definition (with a strict inequality) is originally due to Harsanyi and Selten (1988). Denote a monomorphic state h such that $h = (x, ..., x)$ by h_x.

If x is a strict Nash equilibrium, then h_x is called a *convention* (Young,1993a, 1998). If x is risk (strictly Pareto) dominant, I say that the convention h_x is risk (strictly Pareto) dominant. From Theorem 2, the following corollary immediately follows.

Corollary 1 *In 2×2 Coordination Games with a strictly Pareto-dominant equilibrium, for a sufficiently large sample size, and sufficiently incomplete information, the strictly Pareto-dominant convention is a unique stochastically stable state.*

For certain payoffs, the sample size must to be very large for this result to hold. Consider the game in Figure 3.3, where the equilibrium (A, a) strictly Pareto dominates the equilibrium (B, b), but where (B, b) is the unique risk-dominant equilibrium.

	a	b
A	$2,2$	$-1000,0$
B	$0,-1000$	$1,1$

FIGURE 3.3.

If the process is in state $h_{(B,b)}$ and the sample size $s > 1$, then two simultaneous mistakes, followed by a sequence of imitators, are required to reach $h_{(A,a)}$. One mistake is sufficient to make the reverse transition if this makes the expected payoff to playing strategy B higher than that of A for a better or best replier, or formally if

$$\frac{2(s-1)-1000}{s} \leq \frac{1}{s} \Leftrightarrow s \leq \frac{1003}{2}. \tag{3.4}$$

Hence, only the risk-dominant equilibrium is selected, in the sense that the corresponding convention is stochastically stable, for s such that $1 < s \leq 501$. In order to ensure that only the strictly Pareto-dominant equilibrium is selected, the sample size must be so large that

$$\frac{2(s-2)-2000}{s} > \frac{2}{s} \Leftrightarrow s > 1003. \tag{3.5}$$

For a sample size in between these values, i.e. s such that $501 < s \leq 1003$, both equilibria are selected.

More generally, define the probabilities q_A, q_a, and $q_{\min}$ by

$$q_A = \frac{\pi_2(B,b) - \pi_2(B,a)}{\pi_2(A,a) - \pi_2(A,b) + \pi_2(B,b) - \pi_2(B,a)}, \tag{3.6}$$

$$q_a = \frac{\pi_1(B,b) - \pi_1(A,b)}{\pi_1(A,a) - \pi_1(B,a) + \pi_1(B,b) - \pi_1(A,b)}, \tag{3.7}$$

and

$$q_{\min} = \min\{q_A, q_a, 1 - q_A, 1 - q_a\}. \tag{3.8}$$

In other words, q_A is the probability of strategy A which makes the expected payoff to strategies a and b identical for player 2, and q_a is the probability of strategy A which makes the expected payoff to strategies A and B identical for player 1. $q_{\min}$ is the lowest probability required to make a player indifferent between her pure strategies. Note that in symmetric Coordination Games $q_A = q_a$ and that (see Young,1998) an equilibrium (A, a) is risk dominant if and only if

$$\min\{q_A, q_a\} \leq \min\{1 - q_A, 1 - q_a\}. \tag{3.9}$$

Let Z_{++} be the set of positive integers and let $\lceil y \rceil$ denote the smallest integer greater than or equal to y for any real y. Further, define the function $s_{diff} : Z_{++} \to Z_{++} \cup \{0\}$ and the set $S^{RD} \subset Z_{++}$ as follows:

$$s_{diff}(s) = |\lceil s \min\{q_A, q_a\} \rceil - \lceil s \min\{1 - q_A, 1 - q_a\} \rceil| \tag{3.10}$$

$$S^{RD} = \{s \in Z_{++} : s_{diff}(s) \geq 1 \text{ and } s \leq 1/q_{\min}\}. \tag{3.11}$$

Note that $s_{diff}(s)$ is increasing in s, implying that if there are integer sample sizes s, s', and s'' such that $s < s' < s''$, and $s, s'' \in S^{RD}$, then also $s' \in S^{RD}$.

Theorem 3 *Let Γ be a 2×2 Coordination Game and assume that information is sufficiently incomplete $(s/m \leq 1/2)$.*
i) From any initial state, the unperturbed process $P^{*,m,s,0}$ converges with probability one to a convention and locks in.
ii) For sufficiently large s the stochastically stable states of the perturbed process $P^{*,m,s,\varepsilon}$ correspond one to one with the conventions that are not strictly Pareto dominated.
iii) If there is a unique risk-dominant equilibrium and the set S^{RD} is non-empty, then, for sample size $s \in S^{RD}$, only the risk-dominant convention is stochastically stable.
iv) If both equilibria are risk dominant, then, for sample size $s \in [1, 2/q_{\min}]$, both conventions are stochastically stable.

PROOF: See the Appendix.

Theorem 3 gives sufficient conditions for the selection of different equilibria. The intuition behind the proof of this theorem is the same as in the above example. If the process is in state $h_{(A,a)}$, and (B,b) is a risk-dominant equilibrium, then, for sample sizes such that $q_{\min} \leq 1/s$, or equivalently $s \leq 1/q_{\min}$, only one mistake is necessary to make a sequence of subsequent better or best repliers switch to the risk-dominant pure strategy such that $h_{(B,b)}$ is reached. Second, if the process is in state $h_{(B,b)}$ and (A,a) is not strictly Pareto dominated, then two simultaneous mistakes, followed by a sequence of imitators in one of the populations, are sufficient to reach the basin of attraction of $h_{(A,a)}$. The condition that $s_{diff}(s) \geq 1$ implies that the transition from $h_{(B,b)}$ to $h_{(A,a)}$ cannot be made with only one mistake followed by a sequence of best or better repliers.

The following corollary shows that in symmetric Coordination Games, there always exists a range of sample sizes such that the risk-dominant equilibrium is selected.

Corollary 2 *In symmetric 2×2 Coordination Games, if information is sufficiently incomplete* $(s/m \leq 1/2)$ *and the sample size sufficiently small* $(2 \leq s \leq 1/q_{\min})$*, only the risk-dominant convention is stochastically stable.*

PROOF: If the symmetric game has two risk-dominant equilibria, the result follows immediately from Theorem 3 iv). If the game has a unique risk-dominant equilibrium, then $q_{\min} < 1/2$. This implies that if the sample size is such that $2 \leq s \leq 1/q_{\min}$, then $sq_{\min} \leq 1$, and $s(1 - q_{\min}) > 1$; hence, the requirement that $s_{diff}(s) \geq 1$ is redundant, and the claim follows from Theorem 3 iii). *Q.E.D.*

Holding the sample and memory fixed in symmetric games with one risk-dominant equilibrium and a different Pareto-dominant equilibrium, Corollary 2 implies that the risk-dominant equilibrium should be observed for payoffs such that $q_{\min}$ is low, possibly both equilibria for payoffs such that $q_{\min}$ is intermediate, and the Pareto-dominant equilibrium for payoffs such that $q_{\min}$ is high. This appears to be in line with experimental evidence from symmetric Coordination Games (see Van Huyck (1997) for a survey), although the absolute difference between the equilibrium payoffs and the off-equilibrium payoffs also seems to matter. Given the limited number of repetitions in such experiments, one may naturally question whether the outcome observed actually corresponds to the stochastically stable state.

3.5 Conclusion

In this paper, I analyze stochastic adaptation in finite n-player games played by heterogeneous populations of myopic best repliers, better repliers and imitators. I show that for sufficiently incomplete information, and independently of the population shares, the recurrent sets of the resulting unperturbed Markov chain correspond one to one with the minimal closed sets closed under better replies of the game. Such sets are also asymptotically stable under a large class of deterministic continuous time selection dynamics, containing the replicator dynamics.

The stochastically stable states are contained in the recurrent sets with minimum stochastic potential. In games where one minimal closed set under better replies is Pareto efficient, the span of the stochastically stable states is equal to this set. This result requires a sufficiently large sample size, determined by the payoffs of the game. In all symmetric Coordination Games and many asymmetric Coordination Games, the Pareto efficient equilibrium is selected for a sufficiently large sample size, and the risk-dominant equilibrium is selected for a sufficiently small sample size.

The main contribution of this paper is to analyze the long-run outcome when multiple rules of adaptation, previously analyzed only in isolation, are present in the populations. It is shown that the support of the limiting distribution is independent of the population shares. However, in many games, the exact shape of the limiting distribution will depend on the population shares using the different learning rules, and so will the expected payoff to the individuals employing the various rules. A next step is to study how the individuals using the different learning rules fare for different population shares, and to ask if there exists a rule, or a particular combination of rules, which is evolutionarily stable.

3.6 Appendix

Definition 3 *(Young, 1998) P^{ε} is a* ***regular perturbed Markov process*** *if P^{ε} is irreducible for every $\varepsilon \in (0, \varepsilon^*]$, and for every $h, h' \in H$, $P^{\varepsilon}_{hh'}$ approaches $P^0_{hh'}$ at an exponential rate, i.e. $\lim_{\varepsilon \to 0} P^{\varepsilon}_{hh'} = P^0_{hh'}$ and if $P^{\varepsilon}_{hh'} > 0$ for some $\varepsilon > 0$, then $0 < \lim_{\varepsilon \to 0} \frac{P^{\varepsilon}_{hh'}}{\varepsilon^{r_{hh'}}} < \infty$ for some $r_{hh'} \geq 0$.*

Lemma 2 *$P^{*,m,s,\varepsilon}$ is a regular perturbed Markov process.*

PROOF: The proof of Lemma 2 is completely analogous to the proof in Young (1998, p. 55) *Q.E.D.*

Theorem 4 *(Young, 1998) Let P^{ε} be a regular perturbed Markov process and let μ^{ε} be the unique stationary distribution of P^{ε} for $\varepsilon > 0$. Then $\lim_{\varepsilon \to 0} \mu^{\varepsilon} = \mu^0$ exists, and μ^0 is a stationary distribution of P^0. The stochastically stable states are precisely those states that are contained in the recurrent classes of P^0 having minimum stochastic potential.*

PROOF OF THEOREM 1: Part i) of Theorem 1 will be proved in two steps. In step A, I will prove that from any initial state, $P^{*,m,s,0}$ converges to a minimal γ-configuration. In step B, I will prove that for any MCUBR set, there exists a minimal γ-configuration.

A. Let *γ-adaptive play* be the induced Markov chain when the share of better repliers in each population is one. According to Theorem 1 in Josephson (2000), for a sufficiently large sample size and sufficiently incomplete information ($s/m \leq \frac{1}{|X|}$), the span of each recurrent class of the γ-adaptive play corresponds one to one with an MCUBR set. Since, in each period, there is a positive probability that only better repliers will be selected to play, there is also a positive probability that the Markov chain, from *any* initial state, and in a finite number of periods, will end up in a state h which belongs to a recurrent class E^{γ}_{Z} of the better-reply dynamic, and thus only involves play of pure-strategy profiles in a corresponding MCUBR set Y. From Lemma 1, it follows that the process will never play a pure strategy outside Y, once it has reached such a state. If E^{γ}_{Z} is singleton, this means that it is a recurrent class also under heterogeneous adaptive play $P^{*,m,s,0}$. If E^{γ}_{Z} is not singleton, there is a positive probability that the process makes the transition from h to any other state h' of E^{γ}_{Z} in a finite number of periods and, naturally, without playing any pure strategy outside Y. This follows since in each period, there is a positive probability that only better repliers will be selected to play. Hence, there exists a recurrent class E^{*}_{Y} of $P^{*,m,s,0}$, such that $E^{\gamma}_{Z} \subseteq E^{*}_{Y}$ and $S\left(E^{*}_{Y}\right) \subseteq Y = S\left(E^{\gamma}_{Z}\right)$.

B. Conversely, I will prove that for any MCUBR set, there exists a minimal γ-configuration. If Y is an MCUBR set, then, by Theorem 1 in Josephson (2000), there exists a unique recurrent class E_Z^γ of γ-adaptive play such that $S\left(E_Z^\gamma\right) = Y$. If the Markov chain $P^{*,m,s,0}$ is in a state involving no pure strategy outside Y, there is a positive probability that it will reach a state which belongs to E_Z^γ in a finite number of periods. This follows since $P^{*,m,s,0}$, by A., from any initial state and in a finite number of periods will end up in a state h, which belongs to a recurrent class of the better-reply dynamic, and since, by Lemma 1, the process will never play a pure strategy outside Y, once it has reached a state only involving pure-strategy profiles of Y. It thereafter follows from A. that there exists a recurrent set E_Y^* of $P^{*,m,s,0}$, such that $S(E_Y^*) = Y$.

Part ii) of Theorem 1 follows directly from Theorem 4 in the Appendix since $P^{*,m,s,\varepsilon}$, by Lemma 2 in the Appendix, is a regular perturbed Markov process. *Q.E.D.*

PROOF OF THEOREM 2: Theorem 2 will be proved in three steps. In step A, I show that the transition from any recurrent class to a Pareto-dominant minimal γ-configuration can always be made with at most n mistakes. In step B, I prove that for a sufficiently large sample size, the transition from a Pareto-dominant minimal γ-configuration to any other recurrent class requires at least $n+1$ mistakes. In step C, I use A and B to prove that the Pareto-dominant minimal γ-configuration must have minimum stochastic potential.

A. Assume that the sample size is sufficiently large and $s/m \leq 1/\left|X\right|$ so that, by Theorem 1, the span of each recurrent class of $P^{*,m,s,0}$ corresponds one to one with an MCUBR set. Let $Y \subset X$ be a strictly Pareto-dominant MCUBR set, and let E_Y^* be the corresponding minimal γ-configuration. Assume there exists at least one other recurrent class (otherwise Theorem 2 holds trivially). The transition from such a recurrent class to E_Y^* can always be made with a probability of the order ε^N (or higher). This is, for instance, the case if the individuals in all roles experiment and play a pure-strategy profile $y^t \in Y$ in period t, and a sequence of $m-1$ imitators thereafter are drawn to play in all roles, and all sample y^t.

B. Let E_Z^* be an arbitrary recurrent class, different from E_Y^*. I claim that for a sufficiently large sample size, the probability of the transition from E_Y^* to E_Z^* is at least of the order ε^{N+1}. To make a best or better replier in role i, switch to a pure strategy $x_i \notin Y_i$ after at most n mistakes; the expected payoff to that pure strategy must be greater than for any pure strategy $y_i \in Y_i$,

$$\frac{s-n}{s}u_i(y_i, p^Y_{-i}) + \frac{n}{s}u_i(y_i, p^X_{-i}) \leq \frac{s-n}{s}u_i(x_i, p^Y_{-i}) + \frac{n}{s}u_i(x_i, p^X_{-i}) \tag{3.12}$$

$$\Leftrightarrow$$

$$s \leq \frac{n\left[u_i(y_i, p^Y_{-i}) - u_i(y_i, p^X_{-i}) + u_i(x_i, p^X_{-i}) - u_i(x_i, p^Y_{-i})\right]}{u_i(y_i, p^Y_{-i}) - u_i(x_i, p^Y_{-i})}, \tag{3.13}$$

where $p^Y_{-i} \in \Box(Y_{-i})$, and $p^X_{-i} \in \Box(X_{-i})$. By the boundedness of payoffs and the strict Pareto dominance of Y, the right-hand side of this inequality is clearly bounded for any i, $y_i \in Y_i$, $x_i \notin Y_i$, $p^Y_{-i} \in \Box(Y_{-i})$, and $p^X_{-i} \in \Box(X_{-i})$. Hence, there exists some finite $\hat{s}$, such that for $s > \hat{s}$, strictly more than n mistakes are necessary for a better or best replier to play a pure strategy outside Y.

Similarly, in order to make an imitator maximizing the average realized payoff switch to a pure strategy $x_i \notin Y_i$ after at most n mistakes,

$$\frac{(s-n)\,u_i(y_i, p^Y_{-i}) + (n-1)\,u_i(y_i, p^X_{-i})}{s-1} \leq \pi_i(x_i, x_{-i}) \tag{3.14}$$

$$\Leftrightarrow$$

$$s \leq \frac{n\left[u_i(y_i, p^Y_{-i}) - u_i(y_i, p^X_{-i})\right] + u_i(y_i, p^X_{-i}) - \pi_i(x_i, x_{-i})}{u_i(y_i, p^Y_{-i}) - \pi_i(x_i, x_{-i})}. \tag{3.15}$$

Once again, by the boundedness of payoffs and the strict Pareto dominance of Y, the right-hand side of the last inequality is bounded for any i, $y_i \in Y_i$, $x_i \notin Y_i$, $x_{-i} \in X_{-i}$, $p^Y_{-i} \in \Box(Y_{-i})$, and $p^X_{-i} \in \Box(X_{-i})$. Hence, there exists some finite $\check{s}$, such that if $s > \check{s}$, strictly more than n mistakes are necessary for a better or best replier to play a pure strategy outside Y. An imitator who simply pick the pure strategy with the maximum realized payoff can, of course, never switch to a pure strategy $x_i \notin Y_i$ with less than s mistakes. Thus, for $s > \max\{\hat{s}, \check{s}, N\}$, the resistance of the transition from E^*_Y to E^*_Z must be greater than N.

C. Consider the minimum resistance tree rooted at an arbitrary recurrent class E^*_D, different from E^*_Y. In this tree, there must be a directed edge from E^*_Y to some other recurrent class E^*_Z (possibly identical to E^*_D). Assume that the sample size is so large that the resistance of the transition from E^*_Y to E^*_Z is greater than N (this is possible by B.), and that the stochastic potential of E^*_D is smaller than or equal to that of E^*_Y. Create a new tree by deleting the edge from E^*_Y to E^*_Z, and adding an edge from E^*_D to E^*_Y. The resistance of the deleted edge is, by assumption, greater than N, and the resistance of the added edge is (by A.) smaller than or equal to N. Hence, the total resistance of the new tree is less than that of the tree rooted at E_D^*, contradicting the assumption that the stochastic potential of E_D^* is smaller than or equal to that of E_Y^*. This proves that the stochastic potential of the Pareto-dominant recurrent class E_Y^* is lower than for any other recurrent class and, by Lemma 2 and Theorem 4 in the Appendix, Theorem 2 follows. *Q.E.D.*

PROOF OF THEOREM 3: i) Convergence with probability one to a convention. By Theorem 2 in Josephson (2000), if the share of better repliers is one in all populations and $s/m \leq 1/2$, then the recurrent classes of the unperturbed process are the two monomorphic states $h_{(A,a)}$ and $h_{(B,b)}$. It is clear that these states are absorbing also when the shares of imitators and best repliers are positive. Moreover, since in each period, there is a positive probability that only better repliers are drawn to play, there is a positive probability of reaching one of these two states in a finite number of periods from any other state. Hence, for $s/m \leq 1/2$, the recurrent classes of $P^{*,m,s,0}$ are two states $h_{(A,a)}$ and $h_{(B,b)}$.

ii) Selection of equilibria that are not strictly Pareto dominated. Without loss of generality, assume that (A, a) is a strictly Pareto-dominant equilibrium. Then, the transition from $h_{(B,b)}$ to the basin of attraction of $h_{(A,a)}$ can always be made with two simultaneous mistakes in period t followed by a sequence of $s-1$ imitators in both populations, who all sample x_i^t or more recent strategy-tuples. The reverse transition requires at least three mistakes if the following two conditions are fulfilled. First, in state $h_{(A,a)}$, best or better repliers in one of the populations should not be able to switch to strategy B when the sample contains less then three mistakes by the other population. This is prevented if $q_{\min} > 2/s$, or equivalently if $s > 2/q_{\min}$. Second, imitators in population C_1 should not be able to switch to strategy B with less than three mistakes. They can only do this if two mistakes, one after the other by different populations, make the average payoff to strategy B at least as large as that of A, or formally, if

$$\frac{(s-2)\pi_1(A,a)+\pi_1(A,b)}{s-1} \leq \pi_1(B,a) \tag{3.16}$$

$$\Leftrightarrow$$

$$s \leq \frac{2\pi_1(A,a)-\pi_1(B,a)-\pi_1(A,b)}{\pi_1(A,a)-\pi_i(B,a).} = \hat{s}_1. \tag{3.17}$$

It is clear that for $s > \hat{s}_1$ this inequality does not hold, and a similar critical sample size $\hat{s}_2$ can be computed for individuals in population C_2. Hence, for $s > \max\{\hat{s}_1, \hat{s}_2, 2/q_{\min}\}$, the transition to (B, b) requires at least three mistakes, whereas the reverse transition requires exactly two mistakes (if all imitators choose the pure strategy with maximum sample payoff, i.e. nobody chooses the pure strategy with the maximum *average* payoff in the sample, then it is sufficient that $s > 2/q_{\min}$).

If neither of the equilibria are Pareto dominated, then either $\pi_1(A,a) \geq \pi_1(B,b)$ and $\pi_2(A,a) \leq \pi_2(B,b)$, or $\pi_1(A,a) \leq \pi_1(B,b)$ and $\pi_2(A,a) \geq \pi_2(B,b)$. Without loss of generality, assume that the first two inequalities hold. Then, the transition from $h_{(B,b)}$ to the basin of attraction of $h_{(A,a)}$ can always be made with two simultaneous mistakes in period t, followed by a sequence of $s-1$ imitators in population C_1, who all sample x_i^t and earlier strategy realizations. In an analogous fashion, the transition from $h_{(A,a)}$ to $h_{(B,b)}$ can be made with two mistakes. If, in addition, $s > 1/q_{\min}$, then no better or best repliers are able to switch strategies in state $h_{(A,a)}$ or $h_{(B,b)}$ after only only mistake. Thus, for general Coordination Games and sample size such that $s > \max\{\hat{s}_1, \hat{s}_2, 2/q_{\min}\}$, the stochastic potential is lowest for the state(s) which is (are) not strictly Pareto dominated.

iii) Unique risk-dominant equilibrium. The transition from $h_{(B,b)}$ to the basin of attraction of $h_{(A,a)}$ can be made with k mistakes if individuals in one of the populations, say population C_1, by mistake plays A k times in a row, and $k/s \geq q_A$. This follows since there is a positive probability that better or best repliers are drawn to play in the other population for the next s periods, and that these individuals all sample the k mistakes. Without loss of generality, assume that (A,a) is a risk-dominant equilibrium. Then, the transition from $h_{(B,b)}$ to $h_{(A,a)}$ requires only one mistake if $q_{\min} \leq 1/s$. The requirement that $s_{diff}(s) > 1$ ensures that the reverse transition requires at least two mistakes.

iv) Two risk-dominant equilibria. If both equilibria are risk dominant, then

$$\min\{q_A, q_a\} = \min\{1-q_A, 1-q_a\} = q_{\min}. \tag{3.18}$$

This implies that for $sq_{\min} \leq 2$, the transition from $h_{(A,a)}$ to $h_{(B,b)}$, and the reverse transition, requires the same number of mistakes, $k \leq 2$. Since at least two mistakes are required to make an imitator switch strategies when the process is in a convention, the stochastic potential of the conventions will equal the number of mistakes required to make better and best repliers switch strategy for $s \in [1, 2/q_{\min}]$.

Theorem 3 now follows by Lemma 2 and Theorem 4. *Q.E.D.*

Bibliography

[1] Van Huyck, J. (1997): "Emergent Conventions in Evolutionary Games," to appear in C. Plott and V. Smith, eds., *Handbook of Results in Experimental Economics,* Amsterdam: North-Holland, 2001.

[2] Camerer, C., and T-H. Ho (1999): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica* 67, 827-874.

[3] Foster, D., and P. Young (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* 38, 219-32.

[4] Freidlin, M., and A. Wentzell (1984): *Random Perturbations of Dynamical Systems.* Springer-Verlag.

[5] Fudenberg, D., and D. Levine (1998): *The Theory of Learning in Games.* The MIT Press.

[6] Harsanyi, J., and R. Selten (1988): *A General Theory of Equilibrium Selection in Games.* The MIT Press.

[7] Josephson, J., and A. Matros (2000): "Stochastic Imitation in Finite Games," SSE/EFI Working Paper Series in Economics and Finance No. 363, Stockholm.

[8] Josephson, J. (2000): "Stochastic Better-Reply Dynamics in Games," SSE/EFI Working Paper Series in Economics and Finance No. 400, Stockholm.

[9] Kaarbøe, O. M., and A. F. Tieman (1999): "Equilibrium Selection under Different Learning Modes in Supermodular Games," Discussion paper nr. TI 99-061/1, Tinbergen Institute, Amsterdam.

[10] Kaniovski, Y. M., A. V. Kryazhimiskii, and H. P. Young (2000): "Adaptive Dynamics in Games Played by Heterogeneous Populations," *Games and Economic Behavior* 31, 50-96.

[11] Matros, A. "Clever Agents in Adaptive Learning," SSE/EFI Working Paper Series in Economics and Finance No. 403, Stockholm.

[12] Kandori, M., G. Mailath, and R. Rob (1993): "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica* 61, 29-56.

[13] Ritzberger, K., and J. W. Weibull (1995): "Evolutionary Selection in Normal-Form Games," *Econometrica* 63, 1371-1400.

[14] Sáez Martí, M., and J. W. Weibull (1999): "Clever Agents in Young's Evolutionary Bargaining Model," *Journal of Economic Theory* 86, 268-279.

[15] Simon, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69, 99-118.

[16] Stahl, D. O. (2000): "Action-Reinforcement Learning Versus Rule Learning," Mimeo, University of Texas.

[17] Young, P. H. (1993a): "The Evolution of Conventions," *Econometrica* 61, 57-84.

[18] —— (1993b): "An Evolutionary Model of Bargaining," *Journal of Economic Theory* 59, 145-68.

[19] —— (1998): *Individual Strategy and Social Structure.* Princeton University Press.

Chapter 4

A Numerical Analysis of the Evolutionary Stability of Learning Rules

4.1 Introduction

The bounded rationality paradigm is based on the assumption that people learn to play games by using simple rules of adaptation, often referred to as *learning rules.* The objective is generally to predict which strategies are more likely to be observed in the long run, given that players use a specific learning rule. A problem with this setting is that the learning rule is treated as exogenous and no rationale is provided for the particular choice of learning rule. Evolutionary forces are usually only allowed on the level of simple strategies and not on the higher level of learning rules. In this sense, the area is suffering from a certain open-endedness. In this paper, I attempt to take a step towards closing this open-endedness by developing an evolutionary stability criterion for learning rules and applying this criterion to a set of well-known learning rules using Monte Carlo simulations.

More specifically, I ask if there is a rule such that if applied by a homogeneous population of individuals, it cannot be invaded by mutants using a different rule. I call such an uninvadable rule an *evolutionarily stable learning rule* (ESLR). This concept is an extension of the classical definition of evolutionarily stable strategies (Maynard Smith and Price (1973), Maynard Smith (1974, 1982)) to learning rules and dynamic strategies.

The setting is a world where the members of a large population, consisting of an even number of individuals, in each of a finite number of periods are all randomly matched in pairs to play a finite two-player game.[1] Each individual uses a learning rule, which is a function of his private history of past play, and fitness is measured in terms of expected average payoff. This framework provides a rationale for the use of learning rules and it is of particular interest since very little analysis of learning in this "repeated rematching" context has previously been done.

Technically, learning rules are mappings from the history of past play to the set of pure or mixed strategies. There are many models of learning and I therefore restrict the numerical analysis to a class of learning rules that can be described by the general parametric model of Camerer and Ho (1999), called *experience-weighted attraction learning* (EWA). The rules in this class have experimental support and perform well in an environment where the game changes from time to time. Moreover, the class contains rules which differ considerably in their use of information. Two of the most well-known learning rules, *reinforcement learning* and *fictitious play* (or *belief learning*), are special cases of this model for specific parameter values.

Reinforcement learning is an important model in the psychological literature on individual learning. It was introduced by Bush and Mosteller (1951) although the principle behind the model, that choices which have led to good outcomes in the past are more likely to be repeated in the future, is due to Thorndike (1898). Under reinforcement learning in games, players assign probability distributions to their available pure strategies. If a pure strategy is employed in a particular period, the probability of the same pure strategy being used in the subsequent period increases as a function of the realized payoff. The model has very low information and rationality requirements in the sense that individuals need not know the strategy realizations of the opponents or the payoffs of the game; all that is necessary is knowledge of player-specific past strategy and payoff realizations.

Fictitious play, or belief learning, is a model where the individuals in each of the roles of a game in every period play a pure strategy that is a best reply to the accumulated empirical distribution of their opponents' play. This means that knowledge of the opponents' strategy realizations and the player's own payoff function is required.

Several different models of both reinforcement and fictitious play have been developed over the years. The ones that can be represented by Camerer and Ho's (1999) model correspond to stochastic versions with exponential

[1] To check the robustness of my results, I have also analyzed a different matching scheme, where the individuals only are randomly matched at the start of the first period, and then continue to play against the same opponent for a finite number of periods. The results from the simulation of this matching scheme are in general consistent with the results for the the matching scheme in this paper.

probabilities.[2] This means that each pure strategy in each period is assigned an *attraction*, which is a function of the attraction in the previous period and the payoff to the particular strategy in the current period.[3] The attractions are then exponentially weighted in order to determine the mixed strategy to be employed in the next period. In the case of reinforcement learning, the attractions only depend on the payoff to the pure strategy actually chosen. In the case of belief learning, the hypothetical payoffs to the pure strategies that were not chosen are of equal importance (this is sometimes referred to as *hypothetical reinforcement*). However, Camerer and Ho's (1999) model also permits intermediate cases where payoffs corresponding to pure strategies that were not chosen are given a weight strictly between zero and one. The weight of such hypothetical payoffs is given by a single parameter, δ.

Camerer and Ho's (1999) model also allows initial attractions of different sizes which, in the case of belief learning, correspond to expected payoffs, given a prior distribution over the opponents' pure strategies. I depart from this assumption and set all initial attractions to zero, such that the individuals have almost no prior knowledge of the game they are drawn to play. This implies that the numerical analysis in this paper boils down to testing if any particular value δ corresponds to an ESLR.

In order to test the stability of learning rules, I simulate a large number of outcomes when all members of a finite population with a large share of incumbents and a small share of mutants are randomly matched for a finite number of periods. I then calculate the average payoff for each share of the population. I consider four different games: Prisoners' Dilemma, Coordination, Hawk-Dove and Rock-Scissors-Paper. The main findings are:

- In almost all cases, the learning rule with full hypothetical reinforcement is an ESLR, whereas the learning rule with no hypothetical reinforcement is unstable.

- In the two games with no symmetric pure Nash equilibria – the Hawk-Dove and Rock-Scissors-Paper Games – the results depend on the level of *payoff sensitivity* of the learning rules. This is a parameter of the EWA model determining to what extent differences in attractions for the pure strategies should translate into differences in probabilities. For low payoff sensitivity, several rules appear to be stable, whereas for

[2] Fudenberg and Levine (1998) show that stochastic fictitious play can be derived by maximizing expected payoff given an empirical distribution of the opponents' past play when payoffs are subject to noise.

[3] The term "attraction" is used to make the terminology in this paper consistent with that in Camerer and Ho (1999). This term should not be interpreted in the mathematical sense, but as the weight assigned to a particular strategy.

high payoff sensitivity, only belief learning is stable. The latter finding is, in part, due to that reinforcement learners with a high level of payoff sensitivity quickly become absorbed by a pure strategy, whereas belief learners with the same level of payoff sensitivity continue to adjust their mixed strategies until the last period.

- In 2×2 Coordination Games, the results also depend on the equilibrium payoffs. In such games, belief learning is generally a unique ESLR, but if the ratio of equilibrium payoffs becomes sufficiently small and payoff sensitivity is low, then there are also other stable learning rules.

4.1.1 Related Literature

The present paper is related to the theoretical literature on learning, but also to experimental tests of different learning rules. An early theoretical reference, asking similar questions, is Harley (1981). He analyzes the evolution of learning rules in the context of games with a unique evolutionarily stable strategy (ESS). He assumes the existence of an ESLR and then discusses the properties of such a rule. Harley claims that, given certain assumptions, "...the evolutionarily stable learning rule is a rule for learning evolutionarily stable strategies." He also develops an approximation to such a rule and simulates its behavior in a homogeneous population. The current paper differs from that of Harley (1981) in that it explicitly formulates an evolutionary criterion for learning rules and does not assume the existence of an ESLR. Moreover, the analysis is not limited to games with a single ESS.

Anderlini and Sabourian (1995) develop a dynamic model of the evolution of algorithmic learning rules. They claim that under certain conditions, the frequencies of different learning rules in the population are globally stable and that the limit points of the distribution of strategies correspond to Nash equilibria. However, they do not investigate the properties of the stable learning rules.

Hopkins (2000) investigates the theoretical properties of stochastic fictitious play and perturbed reinforcement learning. The model in this paper is a special case of stochastic fictitious play, when the parameter δ is equal to one, and is similar to Hopkins' version of reinforcement learning when δ is equal to zero. Hopkins finds that the expected motion of both stochastic fictitious play and perturbed reinforcement learning can be written as a perturbed form of the replicator dynamics, and that in many cases, they will therefore have the same asymptotic behavior. In particular, he claims that they have identical local stability properties at mixed equilibria. He also finds that the main difference between the two learning rules is that fictitious play gives rise to faster learning. The analysis in Hopkins (2000) differs from the analysis in this paper, in that it is based on infinite interaction between two players using identical learning rules, but my findings are consistent with Hopkins' (2000) results.

The topic of this paper is also somewhat related to the theoretical literature on evolution in asset markets, such as Blume and Easley (1992, 2000), and Sandroni (2000). In these models, selection operates over beliefs and utility functions and not directly over learning rules, and the authors use a dynamic evolutionary criterion based on wealth accumulation. They find that, under fairly general conditions, correct beliefs are selected for in complete markets, but not necessarily in incomplete markets.

The experimental literature uses a criterion which differs from the evolutionary one introduced in this paper to motivate the use of a particular learning rule. The objective is to find the learning rule which gives the best fit of experimental data. Camerer and Ho (1999) give a concise overview of the most important findings in earlier studies. They argue that the overall picture is unclear, but that comparisons appear to favor reinforcement in constant-sum games and belief learning in Coordination Games. In their own study of asymmetric Constant-Sum Games, Median-Action Games, and Beauty-Contest Games, they find support for a learning rule with parameter values in between reinforcement learning and belief based learning. In particular, they estimate game-specific values of the δ-parameter, which captures the degree of hypothetical reinforcement, strictly between zero and one, and generally around 0.5.

Stahl (2000) compares the prediction performance of seven learning models, including a restricted version of the EWA model. He pools data from a variety of symmetric two-player games and finds a logit best-reply model with inertia and adaptive expectations to perform best, closely followed by the EWA. For the latter, he estimates a value of the δ-parameter of 0.67.

This paper is organized as follows. Section 2 introduces the theoretical model underlying the simulations. Section 3 present the results of the Monte Carlo simulations. Section 4 contains a discussion of the results and Section 5 concludes. Tables and diagrams of some of the simulations can be found in the Appendix.

4.2 Model

Let Γ be a symmetric two-player game on normal form, where each player has a finite pure strategy set $X = \{x^1, ..., x^J\}$, with the mixed-strategy extension $\Delta(X) = \{p \in \Re^J_+ \mid \sum_{j=1}^J p^j = 1\}$. Each player's payoff is represented by the function $\pi : X \times X \to \mathbb{R}$, where $\pi(x, y)$ is the payoff to playing pure strategy x when the opponent is playing pure strategy y. From time to time, all individuals of a finite population, consisting of an even number M of individuals, are drawn to play this game for T periods. The mixed strategy of individual k in period $t \in \{1, 2, ..., T\} = \Upsilon$ is denoted by $p_k(t)$. The pure strategy realization of individual k is denoted by $x_k(t)$ and that of his opponent (in this period) by $y_k(t)$. The sequence

$$h_k(t) = ((x_k(0), y_k(0)), (x_k(1), y_k(1)), ..., (x_k(t-1), y_k(t-1))),$$

where $(x_k(0), y_k(0)) = \emptyset$, is referred to as *individual k's history in period t*. Let $H(t)$ be the finite set of possible such histories at time t, let $H = \cup_{t=1}^{T} H(t)$, and let $\Omega = H(T)$ be the set of outcomes. I define a *learning rule* as a function $f : H \to \Delta(X)$ that maps histories to mixed strategies and denote the set of possible learning rules by $\mathfrak{F}$. Note that according to this definition, initial conditions such as initial history or initial strategy weights are given by the learning rule.

The matching procedure can be described as follows. In each of T periods, *all* members of the population are randomly matched in pairs to play the game Γ against each other. This can be illustrated by an urn with n balls, from which randomly selected pairs of balls (with equal probability) are drawn successively until the urn is empty. This procedure is repeated for a total of T periods, and the draws in each period are independent of the draws in all other periods. Each individual k receives a payoff $\pi(x_k(t), y_k(t))$ in each period and has a private history of realized strategy profiles. The expected payoff for an individual k employing learning rule f in a heterogenous population of size M, where the share of individuals employing rule f is $(1-\varepsilon)$ and the share of individuals employing rule g is ε, is the expected average payoff under the probability measure, $\mu^M_{f,(1-\varepsilon)f+\varepsilon g}$ induced by the two rules present in the population and their respective shares,

$$\begin{aligned} & V^M(f, (1-\varepsilon)f + \varepsilon g) & (4.1) \\ = & \sum_{h_k \in \Omega} \left(\frac{1}{T} \sum_{t=1}^{T} \pi(x_k(t), y_k(t)) \right) \mu^M_{f,(1-\varepsilon)f+\varepsilon g}(h_k) & (4.2) \\ = & E^M_{f,(1-\varepsilon)f+\varepsilon g} \left[\frac{1}{T} \sum_{t=1}^{T} \pi(x_k(t), y_k(t)) \right], & (4.3) \end{aligned}$$

where $\pi(x_k(t), y_k(t))$ refers to the realized payoff to individual k in period t, induced by history h_k.

Let $\mathfrak{F}'$ be an arbitrary non-empty subset of $\mathfrak{F}$. I define the following evolutionary stability criterion for learning rules.

Definition 1 *A learning rule $f \in \mathfrak{F}'$ is* ***evolutionarily stable*** *in the class $\mathfrak{F}'$ if for every $g \in \mathfrak{F}' \setminus f$, there exists an $\hat{\varepsilon}_g > 0$ such that for all $\varepsilon \in (0, \hat{\varepsilon}_g)$,*

$$V^M(f, (1-\varepsilon)f + \varepsilon g) > V^M(g, (1-\varepsilon)f + \varepsilon g). \tag{4.4}$$

4.2.1 Experience Weighted Attraction Learning

In the present paper, I focus on a set of learning rules that can be described by Camerer and Ho's (1999) model of *experienced-weighted attraction* (EWA) learning. These are learning rules such that individual k's probability of strategy x^j in period $t \in \Upsilon$ can be written as

$$p_k^j(t) = \frac{e^{\lambda A_k^j(t-1)}}{\sum_{j=1}^J e^{\lambda A_k^j(t-1)}}, \tag{4.5}$$

where the attraction of strategy x^j is updated according to the formula

$$A_k^j(t) = \frac{\phi N(t-1) A_k^j(t-1) + [\delta + (1-\delta) I(x^j, x_k(t))]\, \pi(x^j, y_k(t))}{N(t)}, \tag{4.6}$$

for $t \in \Upsilon$, and $A_k^j(0)$ is a constant, and where

$$N(t) = \sigma N(t-1) + 1, \tag{4.7}$$

for $t \in \Upsilon$, and $N(0)$ is a constant.[4] $I(x^j, x_k(t))$ is an indicator function which takes the value of one if $x_k(t) = x^j$ and zero otherwise, $y_k(t)$ is the realized pure strategy of the opponent in period t, and ϕ and σ are positive constants.

Note that this class of learning rules includes two of the most common learning rules used in the literature. When $\delta = 0$, $\sigma = \phi$ and $N(0) = \frac{1}{1-\sigma}$, EWA reduces to (average) reinforcement learning.[5] When $\delta = 1$, $\sigma = \phi$ and

$$A_k^j(0) = \sum_{l=1}^J \pi(x^j, y^l) \frac{N_{-k}^l(0)}{\sum_{l=1}^J N_{-k}^l(0)}, \tag{4.8}$$

where $\frac{N_{-k}^l(0)}{\sum_{l=1}^J N_{-k}^l(0)}$ is some initial relative frequency of strategy l, EWA becomes belief learning.

[4] Camerer and Ho (1999) note that it is also possible to model probabilities as a power function of attractors.

[5] Camerer and Ho (1999) distinguishes between *average* and *cumulative* reinforcement, which results if $\rho = 0$ and $N(0) = 1$. The analysis in the present paper is based on average reinforcement.

In order to make the analysis more tractable, I further restrict the set of rules to EWA learning rules such that $\sigma = \phi < 1$, $N(0) = \frac{1}{1-\sigma}$, and

$$A_k^j(0) = 0 \ \forall k, \forall j. \tag{4.9}$$

This means that the initial attractions will not generally correspond to those of belief learning. The assumption of equal initial attractions is motivated by a setting where the players have very limited information about the game before the first period and where they cannot use previous experience.[6] The assumption that $\sigma = \phi$ implies that the discount factor for a belief learner's historical observations of strategy realizations is the same as that for a reinforcement learner's historical attractions. Finally, the value of $N(0)$ corresponds to the steady state value of $N(t)$.[7]

I denote the set of rules with the above parameter values by $\mathfrak{F}_e$. Substituting in (4.6) and (4.7) gives

$$N(t) = \frac{1}{1-\sigma} \text{ for } t \in \Upsilon \tag{4.10}$$

and

$$A_k^j(t) = \sigma A_k^j(t-1) + (1-\sigma)\left[\delta + (1-\delta) I(x^j, x_k(t))\right] \pi(x^j, y_k(t)) \tag{4.11}$$

for $t \in \Upsilon$, and $A_k^j(0) = 0$. The formula in (4.5) now corresponds to belief learning (with modified initial weights) for $\delta = 1$ and to reinforcement learning for $\delta = 0$. The parameter δ captures the extent to which the hypothetical payoffs of pure strategies not played in a period are taken into account. σ is a constant determining the relative weights of recent and historical payoffs in the updating of mixed strategies.

4.3 Numerical Analysis

The analysis is based on Monte Carlo simulations of repeated encounters between individuals using different learning rules (i.e. with different values of δ) belonging to the set $\mathfrak{F}_e$. I focus on four types of games, Prisoners' Dilemma, 2×2 Coordination, Hawk-Dove, and Rock-Scissors-Paper Games. I generally

[6] Although the game is fixed in the below analysis, a rationale for the assumption of uniform initial weight could be a setting where the game is drawn at random from some set of games before the first round of play.

[7] Stahl (2000) finds that a time varying $N(t)$ only improves the predictive power of the model marginally and assumes $N(t) = 1$ for all t. He also assumes all initial attractors to be zero and uses the updating formula to determine the attractors in period one.

set the payoff sensitivity parameter λ in (4.5) to either 1 or 10, σ equal to 0.95 and I assume that δ is an element of the set $D = \{0, 0.25, 0.5, 0.75, 1\}$, but I also test the robustness of my results by trying other parameter values (see the Appendix for a list of simulations).

In the simulations, each member of a population of 100 individuals, among which 10 are mutants with a different learning rule, is randomly matched with another member every period for $T = 100$ periods. The expected payoff to a learning rule is estimated by computing the mean of the average payoff for all individuals with the same learning rule in the population and by simulating 1000 such T-period outcomes. Since the mean payoff difference in each simulation is independently and identically distributed relative to the mean payoff difference in another simulation with the same population mixture, the Central Limit Theorem applies and the mean payoff difference is approximately normally distributed. For each value of δ, the null hypothesis is that the corresponding learning rule is an ESLR. This hypothesis is rejected if the mean payoff to any mutant rule is statistically significantly higher than the mean payoff to the incumbent rule in the class, in accordance with Definition 1 above. More specifically, the decision rule is as follows. The null hypothesis,

$$H_0^\delta : f_\delta \text{ is an ESLR in the class } \mathfrak{F}_e,$$

is rejected in favor of the alternative hypothesis,

$$H_1^\delta : f_\delta \text{ is not an ESLR in the class } \mathfrak{F}_e,$$

if and only if, for some $\delta' \in D \setminus \delta$,

$$\frac{\hat{V}(f_\delta, (1-\varepsilon)f_\delta + \varepsilon f_{\delta'}) - \hat{V}(f_{\delta'}, (1-\varepsilon)f_\delta + \varepsilon f_{\delta'})}{\frac{1}{\sqrt{T}} s_\Delta \left[(1-\varepsilon)f_\delta + \varepsilon f_{\delta'}\right]} < -z_\alpha, \tag{4.12}$$

where $\hat{V}$ is the estimated average payoff, $s_\Delta \left[(1-\varepsilon)f_\delta + \varepsilon f_{\delta'}\right]$ denotes the sample standard deviation of the difference in mean average payoffs, computed over the 1000 simulations, and z_α is the critical value of the standard normal distribution.

Prisoners' Dilemma Games

Table 4.1 depicts the mean of the average payoffs among 90 incumbents, with a δ given in the left-most column, and 10 mutants, with a δ given in the top row, playing the game in Figure 4.1 for 1000×100 periods, when payoff sensitivity $\lambda = 10$. The value in brackets corresponds to the z-statistic of the differences in means, i.e. the difference, computed as the average incumbent payoff minus the average mutant payoff, divided by the standard error of the difference. As explained above, the null hypothesis that a learning rule with a particular δ is an ESLR can be rejected if this value is smaller than the critical value of the standard normal distribution, $-z_\alpha$ for some mutant learning rule in the class, different from f_δ.

It follows from the table that the null can be rejected for all learning rules except the one with $\delta = 1$ at the 10% and 5% significance level ($z_{0.05} = 1.645$, and $z_{0.10} = 1.282$). This is also illustrated by the diagram in Figure 4.3, where the standardized payoff difference (the z-statistic) between incumbent and mutant payoffs is plotted for different values of incumbent and mutant δ. In the diagram, the difference is set to zero for homogenous populations. The result is robust to changes in payoff sensitivity λ, initial conditions, payoff matrix, and the size of the mutant invasion.

The standard deviation of payoffs among learning rules with $\delta = 0$ is considerably larger than for other values of δ. The volatility of payoffs also depends on payoff sensitivity. If λ is reduced from 10 to 1, the range of standard deviations decreases considerably. For the high value of λ, convergence to the Nash equilibrium is fast. For the low value, the population share using the equilibrium strategy increases more slowly and keeps oscillating.

	y^1	y^2
x^1	4	0
x^2	5	3

FIGURE 4.1

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.8977 2.8876 (1.01)	2.8574 3.3005 (-70.20)	2.8570 3.3333 (-65.55)	2.8582 3.3303 (-65.97)	2.8572 3.3315 (-67.70)
0.25	3.0001 2.8610 (33.65)	2.9906 2.9907 (-0.14)	2.9891 3.0106 (-38.93)	2.9886 3.0166 (-54.25)	2.9883 3.0185 (-56.95)
0.50	3.0030 2.9018 (36.79)	2.9973 2.9788 (27.54)	2.9959 2.9962 (-0.88)	2.9955 3.0015 (-19.07)	2.9953 3.0031 (-26.46)
0.75	3.0039 2.9061 (38.71)	2.9987 2.9774 (33.98)	2.9975 2.9918 (14.86)	2.9970 2.9973 (-1.05)	2.9969 2.9988 (-6.85)
1.00	3.0037 2.9126 (39.39)	2.9991 2.9765 (34.83)	2.9979 2.9903 (21.62)	2.9975 2.9952 (8.07)	2.9973 2.9979 (-2.17)

TABLE 4.1–Mean payoffs and standardized payoff differences from playing the game in Figure 4.1.

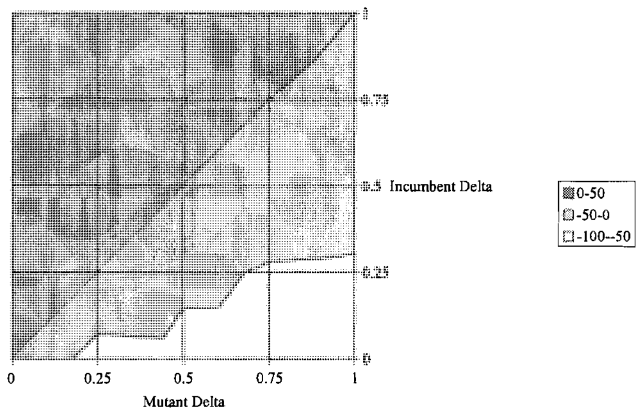


FIGURE 4.2–Standardized payoff difference between incumbents and mutants from Table 4.1.

2×2 Coordination Games

Table 4.2 and Figure 4.4 show the results from the simulations of the Coordination Game in Figure 4.3, with payoff sensitivity $\lambda = 10$. Once again, the null hypothesis can be rejected for all learning rules except $\delta = 1$ at 5% or 10% significance. This result is robust to changes in λ, the size of the invasion, and the initial conditions. However, when the ratio of diagonal payoffs is small ($\pi(x^1, y^1) = 1.1$ instead of 2) *and* $\lambda = 1$, then the null cannot be rejected for any of the rules $\delta = 1, \delta = 0.75$, and $\delta = 0.25$. From the table, it also follows that the outcome for a homogenous population of belief learners Pareto dominates that of a population of reinforcement learners.

	y^1	y^2
x^1	2	0
x^2	0	1

FIGURE 4.3

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.8583 1.8588 (-0.58)	1.8618 1.8770 (-24.61)	1.8640 1.8833 (-36.44)	1.8651 1.8866 (-40.05)	1.8654 1.8868 (-40.00)
0.25	1.8993 1.8866 (17.67)	1.9020 1.9020 (0.12)	1.9022 1.9069 (-11.17)	1.9042 1.9108 (-16.35)	1.9036 1.9113 (-19.32)
0.50	1.9131 1.8973 (23.86)	1.9148 1.9103 (10.45)	1.9157 1.9165 (-2.14)	1.9161 1.9184 (-6.12)	1.9168 1.9207 (-10.23)
0.75	1.9202 1.9022 (28.15)	1.9222 1.9152 (15.89)	1.9227 1.9206 (5.19)	1.9235 1.9239 (-1.17)	1.9240 1.9249 (-2.53)
1.00	1.9257 1.9074 (28.57)	1.9268 1.9194 (16.94)	1.9274 1.9235 (9.89)	1.9280 1.9260 (5.49)	1.9286 1.9290 (-1.28)

TABLE 4.2-Mean payoffs and standardized payoff differences from playing the game in Figure 4.3.

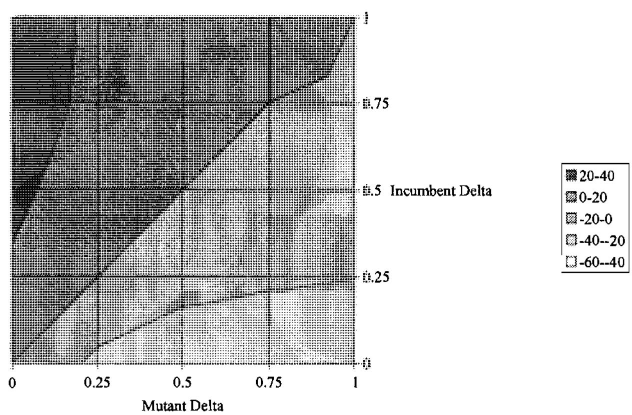


FIGURE 4.4–Standardized payoff difference between incumbents and mutants from Table 4.2.

Hawk-Dove Games

Table 4.3 and Figure 4.6 illustrate the results from the simulations of the game in Figure 4.5. For this game, the results are sensitive to the level of payoff sensitivity λ and the payoff matrix. When $\lambda = 10$, the null hypothesis cannot be rejected for the learning rules with $\delta = 0.75$ and $\delta = 1$ at the 10% level, whereas all other learning rules in the class appear to be unstable. If $\pi(x^1, y^2)$ is increased from 4 to 10, such that the initial uniform distribution is further from the mixed equilibrium, then the null can be rejected for all rules except $\delta = 0.75$.

Part of the reason why learning rules with low δ are not stable and cannot invade other learning rules in this game when payoff sensitivity is high is that they quickly become absorbed by a pure strategy (see Figure 4.12 in the Appendix), something which can be exploited by learning rules with high δ that do not lock in on a particular pure strategy. This tendency for reinforcement learners to be absorbed has previously been noted by Fudenberg and Levine (1998).

For $\lambda = 1$, reinforcement learners no longer lock in on a particular strategy, but oscillate around the mixed equilibrium (see Figure 4.11 in the Appendix), which somewhat reduces the evolutionary advantage of belief-learners. For the game in Figure 4.5, the null hypothesis can be rejected for all rules except $\delta = 0.25$, $\delta = 0.75$, and $\delta = 1.0$ at the 10% level. In the game with $\pi(x^1, y^2) = 10$, the result is unchanged and all rules except $\delta = 0.75$ can be rejected.

The matrix in Table 4.3 also illustrates the potential trade-off between the Pareto efficiency and the evolutionary stability of a learning rule. The learning rule with $\delta = 0$ strictly dominates all other learning rules, but it is not sustainable since, in the case of an invasion, mutants with higher δ earn higher mean payoffs.

	y^1	y^2
x^1	0	4
x^2	1	2

FIGURE 4.5

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.7605 1.7612 (-0.20)	1.7508 1.7843 (-9.23)	1.7100 1.8353 (-36.42)	1.6692 1.8445 (-54.56)	1.6794 1.8397 (-50.94)
0.25	1.6889 1.6657 (8.06)	1.6819 1.6795 (0.82)	1.6518 1.7150 (-22.42)	1.6148 1.7302 (-43.28)	1.6238 1.7249 (-38.13)
0.50	1.5512 1.5206 (14.25)	1.5527 1.5283 (11.29)	1.5420 1.5402 (0.77)	1.5183 1.5505 (-14.54)	1.5256 1.5557 (-14.27)
0.75	1.4512 1.4279 (13.94)	1.4510 1.4306 (11.78)	1.4469 1.4346 (6.69)	1.4418 1.4402 (0.82)	1.4422 1.4397 (1.32)
1.00	1.4604 1.4391 (12.37)	1.4601 1.4397 (11.47)	1.4551 1.4419 (6.99)	1.4469 1.4473 (-0.19)	1.4468 1.4446 (1.21)

TABLE 4.3-Mean payoffs and standardized payoff differences from playing the game in Figure 4.5.

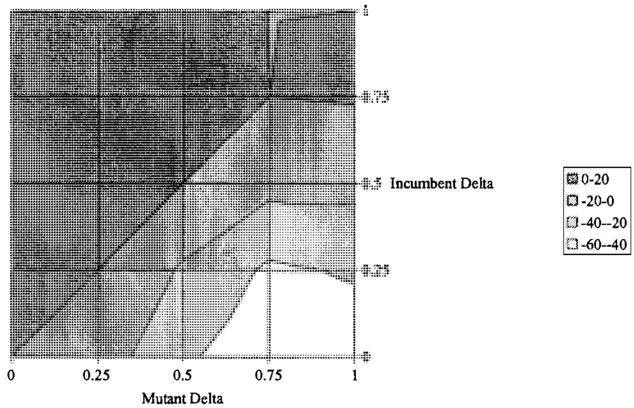


FIGURE 4.6–Standardized payoff difference between incumbents and mutants from Table 4.3.

Rock-Scissors-Paper

In the Rock-Scissors-Paper Game of Figure 4.7, the outcome is sensitive to payoff sensitivity. When the payoff sensitivity is $\lambda = 1$, the null hypothesis can be rejected for all learning rules in the class except $\delta = 0.0$, $\delta = 0.75$, and $\delta = 1.0$. All learning rules oscillate around the mixed equilibrium.

Table 4.4 and Figure 4.8 illustrate the case when $\lambda = 10$. As can be seen, the null hypothesis can be rejected for all learning rules except $\delta = 1$. As in the Haw-Dove game, the instability of rules with low δ for high values of payoff sensitivity can, in part, be explained by their tendency to lock in on a pure strategy at an early stage.

	y^1	y^2	y^3
x^1	1	2	0
x^2	0	1	2
x^3	2	0	1

FIGURE 4.7

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.0001 0.9994 (0.58)	0.9987 1.0121 (-11.14)	0.9969 1.0278 (-23.92)	0.9970 1.0271 (-22.98)	0.9974 1.0236 (-21.35)
0.25	1.0006 0.9945 (5.46)	1.0000 0.9998 (0.25)	0.9984 1.0146 (-15.01)	0.9978 1.0200 (-21.19)	0.9981 1.0175 (-18.65)
0.50	1.0001 0.9988 (1.44)	1.0003 0.9976 (3.01)	0.9999 1.0009 (-1.16)	0.9995 1.0047 (-5.92)	0.9992 1.0071 (-8.89)
0.75	1.0004 0.9967 (4.16)	1.0002 0.9983 (2.24)	1.0002 0.9984 (2.22)	1.0000 1.0004 (-0.50)	0.9997 1.0026 (-3.34)
1.00	1.0003 0.9971 (3.66)	1.0003 0.9971 (3.59)	1.0005 0.9953 (6.06)	1.0002 0.9980 (2.65)	1.0000 0.9998 (0.28)

TABLE 4.4-Mean payoffs and standardized payoff differences from playing the game in Figure 4.4.

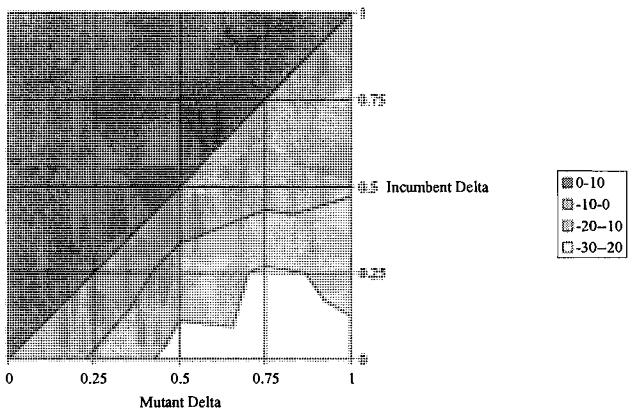


FIGURE 4.8–Standardized payoff difference between incumbents and mutants from Table 4.8.

4.3.1 Summary of Results

Table 4.9 summarizes the results of the simulations. The main finding is that belief learning is the only learning rule which is evolutionarily stable in almost all settings, whereas reinforcement learning is unstable in almost all settings.

In the Hawk-Dove and Rock-Scissors-Paper Games, the results depend on the payoff sensitivity. Learning rules with low degrees of hypothetical reinforcement are highly unstable for high payoff sensitivity. Part of the explanation is that such rules rapidly become absorbed by a pure strategy, whereas belief learners with the same level of payoff sensitivity continue to adjust their mixed strategies until the last period.

In Coordination and Hawk-Dove Games, the results also depend on the equilibrium payoffs. In the Coordination Game, belief learning is generally a unique ESLR, but if the ratio of equilibrium payoffs becomes sufficiently small, then there are also other learning rules for which the null cannot be rejected. Similarly, in the Hawk-Dove Game, belief learning with $\delta = 0.75$ is a unique ESLR for high payoff ratio, but for a smaller ratio it seems that there are also other stable learning rules.

Game	λ	**ESLR at the 10% significance level**
Prisoners' Dilemma	1	$\delta = 1.0$
	10	$\delta = 1.0$
Coordination	1	$\delta = 1.0$, for low payoff ratio $\delta = 0.25, 0.75, 1.0$
	10	$\delta = 1.0$, for low payoff ratio $\delta = 1.0$
Hawk-Dove	1	$\delta = 0.25, 0.75, 1.0$, for high payoff ratio $\delta = 0.75$
	10	$\delta = 0.75, 1.0$, for high payoff ratio $\delta = 0.75$
Rock-Scissors-Paper	1	$\delta = 0.0, 0.75, 1.0$
	10	$\delta = 1.0$

TABLE 4.9-Summary of the results from the different simulations.

4.4 Discussion

Hopkins (2000) investigates the theoretical properties of stochastic fictitious play and perturbed reinforcement learning in a setting where two individuals using identical learning rules interact for an infinite number of periods. He demonstrates that the expected motion of both stochastic fictitious play and perturbed reinforcement learning can be written as a perturbed form of the replicator dynamics, and therefore, in many cases, will have the same asymptotic behavior. In particular, he claims that they have identical local stability properties at mixed equilibria and that the main difference between the two learning rules is that fictitious play gives rise to faster learning. The results in this paper indicate that speed of learning is indeed an important factor in explaining the stability of belief learning and that the difference between rules with high and low degrees of hypothetical reinforcement is smaller in games with mixed equilibria. However, other factors, such as a high probability of convergence to the equilibrium with the highest payoff in 2×2 Coordination Games and a low probability of absorption by a pure strategy in games with no symmetric pure equilibria, also seem important.

Camerer and Ho (2000) estimate separate sets of parameters for asymmetric constant-sum games, median-action games and beauty-contest games. Their estimates of the degree of hypothetical enforcement δ are generally around 0.5, that of the discount factor ϕ in the range of 0.8 to 1.0, that of the second discount factor σ in the range of 0 to ϕ, and the payoff sensitivity λ varies from 0.2 to 18. Reinforcement learning and belief learning are generally rejected in favor of an intermediate model. Stahl (2000) pools data from several symmetric two-player games and estimates a δ of 0.67. Hence, the two studies lend support to the hypothesis that people take hypothetical payoffs into account, but especially the former study seems to find lower degrees of hypothetical reinforcement than predicted by the evolutionary analysis in this paper.

One should, however, be cautious in making a direct comparison with the results in Camerer and Ho (2000). First of all, the games played in their experiments differ considerably from, and are more complex, than the ones analyzed in this paper. Second, the learning rules in this paper do not exactly correspond to theirs. In particular, Camerer and Ho allow learning rules with different initial attractions, whereas I assume that the players give equal weight to all their pure strategies at the start of the first period of play.

The setting in this paper is, at least in some respects, more similar to that in Stahl (2000). He also considers finite symmetric two-player games with and without symmetric pure equilibria. Moreover, he assumes the initial attractions of the EWA model to be zero, and use the updating formula to determine their values in period one.

A final comment concerns the environment where the learning rules operate. Although the game is fixed in this paper, the general idea is to find a learning rule which is evolutionarily stable under various conditions and can survive in a setting where the game changes from time to time – in many ways a more realistic description of human interaction. The results in this paper indicate that belief learning is indeed such a robust rule, although more analysis is needed to confirm this hypothesis.

4.5 Conclusion

In this paper, I define an evolutionary stability criterion for learning rules. I then apply this criterion to a class of rules which contains versions of two of the most well-known learning rules, reinforcement learning and belief learning, as well as intermediate rules in terms of hypothetical reinforcement. I perform Monte Carlo simulations of a matching scheme where all members of a large population are rematched in every period and I find that maximum or close to maximum hypothetical reinforcement is the only learning rule that is evolutionarily stable for almost all the games studied. I also find that evolutionary stability in some games hinges critically on payoff sensitivity and the relative payoffs of the game.

The objective of this paper is to take a step towards closing the open-endedness of the bounded rationality paradigm. A next step might be to apply this analysis to a larger set of learning rules or, more importantly, to obtain theoretical results which can help explain the observations in this paper.

4.6 Appendix

4.6.1 Plots of Simulated Outcomes

The following diagrams illustrate the share of individuals playing strategy x^1 among 90 incumbents, using a learning rule with $\delta = 1$, and 10 mutants, using a learning rule with $\delta = 0$, in a single simulation. Initial attractions are zero for all pure strategies and $\sigma = 0.95$.

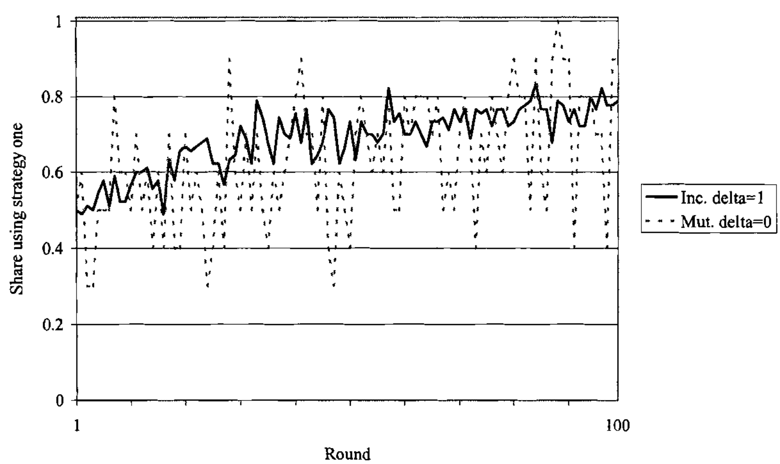


FIGURE 4.9–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 4.3 when $\lambda = 1$.

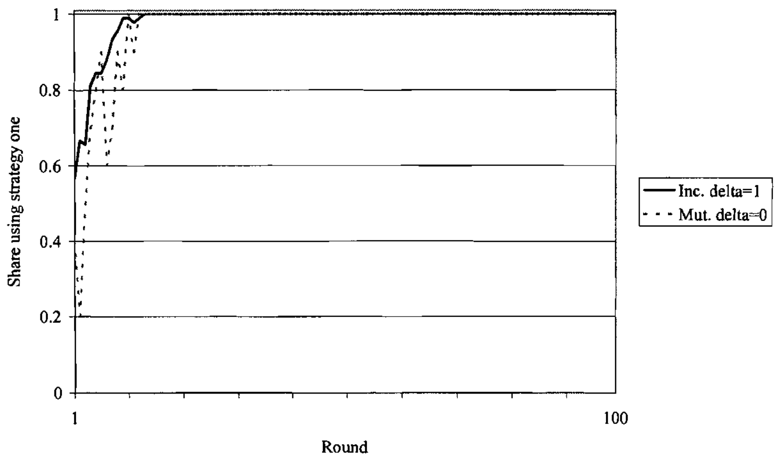


FIGURE 4.10–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 4.3. when $\lambda = 10$.

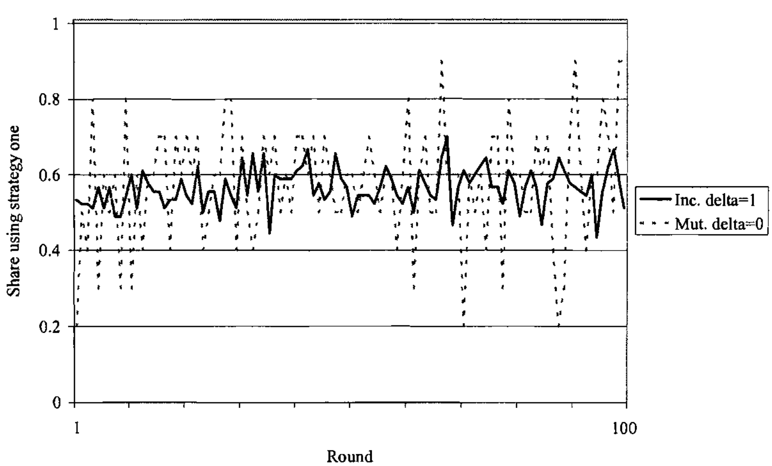


FIGURE 4.11–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in the game in Figure 4.5 when $\lambda = 1$.

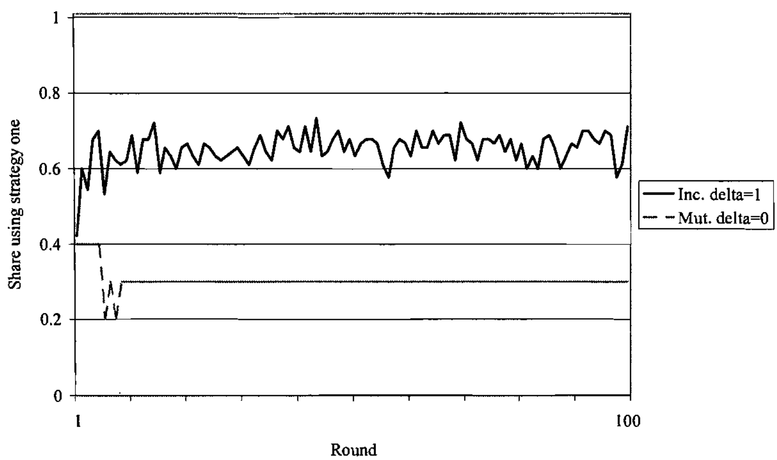


FIGURE 4.12–Share of incumbents (solid) and mutants (dashed) using pure strategy x^1 in in the game in Figure 4.5 when $\lambda = 10$.

4.6.2 Tables of Simulated Payoffs

The following tables report the simulated average payoffs to an incumbent learning rule with a δ given by the left-most column, and an incumbent rule with a δ given by the top row. In each cell, the top-left number is the average payoff for all individuals with the incumbent rule, the top-right number the average payoff for all individuals with the mutant rule, and the figure in brackets is the difference between these numbers divided by the estimated standard error of the difference. In all simulations reported below, a population, consisting of 90 incumbents and 10 mutants, plays the game for 100 periods. Initial attractions are zero for all pure strategies and $\sigma = 0.95$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.9252 2.9239 (0.61)	2.9218 2.9576 (-21.86)	2.9201 2.9696 (-34.76)	2.9200 2.9722 (-39.55)	2.9195 2.9738 (-42.92)
0.25	2.9306 2.8955 (16.77)	2.9280 2.9241 (2.45)	2.9259 2.9425 (-11.59)	2.9252 2.9490 (-18.95)	2.9246 2.9543 (-24.89)
0.50	2.9328 2.8823 (24.80)	2.9298 2.9143 (9.96)	2.9282 2.9275 (0.49)	2.9270 2.9373 (-8.36)	2.9269 2.9410 (-12.20)
0.75	2.9342 2.8711 (33.24)	2.9312 2.9022 (18.88)	2.9296 2.9194 (7.86)	2.9286 2.9287 (-0.10)	2.9282 2.9341 (-5.23)
1.00	2.9346 2.8670 (36.11)	2.9321 2.8940 (24.95)	2.9304 2.9124 (13.73)	2.9295 2.9221 (6.10)	2.9286 2.9280 (0.61)

TABLE 4.10-Mean payoffs and standardized payoff differences from playing the game in Figure 4.1 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.8918 0.8906 (1.07)	0.8958 0.9064 (-9.08)	0.8992 0.9205 (-18.73)	0.9011 0.9293 (-25.21)	0.9015 0.9402 (-34.43)
0.25	0.9310 0.9172 (11.80)	0.9361 0.9341 (1.74)	0.9389 0.9496 (-9.50)	0.9421 0.9627 (-18.15)	0.9439 0.9713 (-25.32)
0.50	0.9692 0.9419 (24.00)	0.9734 0.9623 (10.10)	0.9765 0.9775 (-0.91)	0.9893 0.9896 (-0.26)	0.9838 1.0015 (-15.11)
0.75	1.0066 0.9657 (34.92)	1.0097 0.9868 (20.71)	1.0130 1.0023 (9.74)	1.0170 1.0169 (0.08)	1.0182 1.0269 (-7.83)
1.00	1.0390 0.9854 (45.31)	1.0441 1.0072 (30.63)	1.0464 1.0253 (17.68)	1.0488 1.0394 (8.12)	1.0527 1.0529 (-0.21)

TABLE 4.11-Mean payoffs and standardized payoff differences from playing the game in Figure 4.3 when $\lambda = 1$ and $\varepsilon = 0.1$.

	y^1	y^2
x^1	1.1	0
x^2	0	1

FIGURE 4.13

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.5263 0.5255 (1.45)	0.5262 0.5268 (-1.12)	0.5259 0.5270 (-1.99)	0.5258 0.5266 (-1.42)	0.5264 0.5265 (-0.13)
0.25	0.5262 0.5260 (0.33)	0.5261 0.5250 (2.16)	0.5263 0.5263 (-0.14)	0.5263 0.5257 (1.08)	0.5262 0.5265 (-0.64)
0.50	0.5265 0.5260 (0.90)	0.5264 0.5275 (-2.01)	0.5264 0.5259 (0.97)	0.5265 0.5274 (-1.63)	0.5270 0.5265 (0.87)
0.75	0.5269 0.5266 (0.60)	0.5268 0.5262 (1.15)	0.5274 0.5278 (-0.59)	0.5272 0.5262 (1.73)	0.5274 0.5273 (0.17)
1.00	0.5269 0.5263 (1.02)	0.5271 0.5276 (-1.02)	0.5271 0.5274 (-0.52)	0.5266 0.5261 (0.87)	0.5274 0.5270 (0.78)

TABLE 4.12-Mean payoffs and standardized payoff differences from playing the game in Figure 4.13 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	0.7548 0.7543 (0.48)	0.7686 0.7980 (-34.63)	0.7817 0.8181 (-41.35)	0.7738 0.8117 (-45.47)	0.7757 0.8131 (-46.57)
0.25	0.8888 0.8602 (26.82)	0.8938 0.8936 (0.33)	0.9067 0.9147 (-15.91)	0.9054 0.9159 (-22.81)	0.9054 0.9158 (-22.31)
0.50	0.9302 0.8950 (32.96)	0.9376 0.9308 (12.82)	0.9446 0.9451 (-1.35)	0.9431 0.9457 (-6.86)	0.9415 0.9441 (-6.64)
0.75	0.9473 0.9127 (32.57)	0.9581 0.9493 (17.28)	0.9562 0.9538 (5.90)	0.9632 0.9638 (-1.66)	0.9626 0.9634 (-2.41)
1.00	0.9604 0.9279 (32.69)	0.9656 0.9546 (22.00)	0.9713 0.9680 (8.64)	0.9718 0.9709 (2.56)	0.9727 0.9728 (-0.43)

TABLE 4.13-Mean payoffs and standardized payoff differences from playing the game in Figure 4.13 when $\lambda = 10$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.6079 1.6072 (0.35)	1.6067 1.6074 (-0.37)	1.6082 1.6094 (-0.67)	1.6080 1.6104 (-1.43)	1.6080 1.6111 (-1.80)
0.25	1.6055 1.6053 (0.07)	1.6053 1.6057 (-0.22)	1.6050 1.6061 (-0.62)	1.6060 1.6055 (0.29)	1.6066 1.6057 (0.52)
0.50	1.6073 1.6062 (0.64)	1.6060 1.6076 (-0.90)	1.6067 1.6041 (1.54)	1.6066 1.6079 (-0.73)	1.6067 1.6089 (-1.40)
0.75	1.6097 1.6103 (-0.33)	1.6089 1.6099 (-0.57)	1.6103 1.6070 (1.89)	1.6100 1.6115 (-0.88)	1.6100 1.6088 (0.73)
1.00	1.6124 1.6095 (1.63)	1.6117 1.6134 (-1.00)	1.6123 1.6132 (-0.52)	1.6123 1.6125 (-0.01)	1.6132 1.6124 (0.46)

TABLE 4.14-Mean payoffs and standardized payoff differences from playing the game in Figure 4.5 when $\lambda = 1$ and $\varepsilon = 0.1$.

	y^1	y^2
x^1	0	10
x^2	1	2

FIGURE 4.14

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	2.2135 2.2164 (-0.57)	2.2103 2.2357 (-5.10)	2.2131 2.2311 (-3.71)	2.2150 2.2248 (-2.08)	2.2188 2.2322 (-2.95)
0.25	2.2072 2.1882 (3.51)	2.2026 2.2008 (0.35)	2.2042 2.2062 (-0.41)	2.2044 2.2130 (-1.88)	2.2082 2.2052 (0.63)
0.50	2.2166 2.1936 (4.34)	2.2134 2.2101 (0.66)	2.2130 2.2124 (0.14)	2.2160 2.2138 (0.46)	2.2181 2.2105 (1.65)
0.75	2.2349 2.2071 (5.30)	2.2300 2.2271 (0.56)	2.2290 2.2365 (-1.55)	2.2342 2.2262 (1.65)	2.2362 2.2219 (3.10)
1.00	2.2524 2.2383 (2.72)	2.2485 2.2524 (-0.81)	2.2476 2.2550 (-1.49)	2.2502 2.2627 (-2.72)	2.2542 2.2553 (-0.23)

TABLE 4.15-Mean payoffs and standardized payoff differences from playing the game in Figure 4.14 when $\lambda = 1$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	3.1040 3.0822 (1.34)	2.8984 3.8830 (-83.83)	2.8590 3.9861 (-112.93)	2.8618 4.0014 (-118.06)	2.8655 4.0164 (-119.27)
0.25	2.2468 1.9268 (39.42)	2.1574 2.1666 (-1.32)	2.0830 2.2953 (-37.73)	2.0814 2.3095 (-43.77)	2.0944 2.3237 (-42.27)
0.50	1.7555 1.5211 (47.15)	1.7190 1.6076 (22.07)	1.6824 1.6874 (-1.09)	1.6824 1.6896 (-1.62)	1.6872 1.6863 (0.20)
0.75	1.7327 1.4996 (48.20)	1.7002 1.5665 (28.17)	1.6563 1.6369 (4.29)	1.6492 1.6517 (-0.57)	1.6537 1.6473 (1.52)
1.00	1.7903 1.5412 (44.56)	1.7503 1.6142 (24.55)	1.6979 1.6945 (0.73)	1.6903 1.7068 (-3.69)	1.6966 1.6968 (-0.04)

TABLE 4.16-Mean payoffs and standardized payoff differences from playing the game in Figure 4.14 when $\lambda = 10$ and $\varepsilon = 0.1$.

Delta Mutant / Delta Incumbent	0.00	0.25	0.50	0.75	1.00
0.00	1.0000 1.0002 (-0.22)	1.0001 0.9992 (0.98)	1.0000 1.0002 (-0.19)	1.0000 1.0000 (0.00)	1.0002 0.9983 (2.14)
0.25	0.9998 1.0016 (-2.01)	1.0000 0.9996 (0.57)	1.0000 1.0000 (-0.03)	0.9998 1.0020 (-2.66)	1.0000 1.0004 (-0.53)
0.50	1.0000 0.9999 (0.11)	1.0000 0.9997 (0.44)	1.0000 0.9999 (0.07)	0.9998 1.0019 (-2.39)	1.0000 1.0001 (-0.08)
0.75	1.0001 0.9995 (0.64)	1.0001 0.9988 (1.52)	1.0000 1.0001 (-0.11)	1.0001 0.9994 (0.79)	1.0001 0.9992 (1.14)
1.00	1.0001 0.9995 (0.64)	0.9999 1.0005 (-0.67)	1.0000 1.0000 (0.01)	1.0000 1.0001 (-0.13)	1.0000 0.9998 (0.27)

TABLE 4.17-Mean payoffs and standardized payoff differences from playing the game in Figure 4.7 when $\lambda = 1$ and $\varepsilon = 0.1$.

Bibliography

[1] Anderlini, L., and H. Sabourian (1995): "The Evolution of Algorithmic Learning Rules: A Global Stability Result," Mimeo, Cambridge University.

[2] Blume, L., and D. Easley (1992): "Evolution and Market Behavior," *Journal of Economic Theory*, 58(1), 9-40.

[3] —— (2000): "If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets," Mimeo, Cornell University.

[4] Bush, R. and F. Mosteller (1951): "A Mathematical Model for Simple Learning," *Psychological Review* 58, 313-323.

[5] Camerer, C. and T-H. Ho (1999): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827-874.

[6] Fudenberg, D., and D. Levine (1998): *The Theory of Learning in Games.* The MIT Press.

[7] Harley, C. B. (1981): "Learning the Evolutionarily Stable Strategy," *Journal of Theoretical Biology*, 89, 611-633.

[8] Hopkins, E. (2000): "Two Competing Models of How People Learn in Games," Mimeo, University of Edinburgh.

[9] Maynard Smith, J., and G. R. Price (1973): "The Logic of Animal Conflict," *Nature*, 246, 15-18.

[10] Maynard Smith, J. (1974): "The Theory of Games and the Evolution of Animal Conflicts," *Journal of Theoretical Biology*, 47, 209-221.

[11] Sandroni, I. (2000): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 68, 1303-1341.

[12] Stahl, D. O. (2000): "Action-Reinforcement Learning Versus Rule Learning," Mimeo, University of Texas.

[13] Thorndike, E. L. (1898): "Animal Intelligence: An Experimental Study of the Associative Process in Animals," *Psychological Monographs* 2.

List of Symbols

Γ	Finite game in strategic form
$i \in N$	Player role
C_i	Population in player role i
$x_i \in X_i$	Pure strategy for player role i
$x \in X$	Pure-strategy profile
$p_i(x_i) \in [0,1]$	Player role i's probability mass on the pure strategy x_i
$p_i \in \Delta(X_i)$	Mixed strategy for player role i
$p \in \square(X)$	Mixed-strategy profile
π_i	Payoff function for individuals in population C_i
Π_i	Set of possible payoffs to player i
u_i	Expected payoff function for individuals in population C_i
t	Time period
$h \in H$	History, state
h_x	Monomorphic state, state of the form $h_x = (x, ..., x)$
m	Memory size
s	Sample size
w_i	Strategy sample for player role i
v_i	Payoff sample for player role i
$p_i(x_i \mid h)$	Conditional probability that i chooses strategy x_i in state h
α	Maximum average correspondence
β	Best-reply correspondence
γ	Better-reply correspondence
$P^{\alpha,m,s,0}$	Unperturbed process of imitation play
$P^{\beta,m,s,0}$	Unperturbed process of β-adaptive play
$P^{\gamma,m,s,0}$	Unperturbed process of γ-adaptive play
$P^{*,m,s,0}$	Unperturbed process of heterogeneous adaptive play
$P_{hh'}^{\cdot,m,s,0}$	Transition probability from h to h'
ε	Probability of random choice
$P^{\cdot,m,s,\varepsilon}$	Perturbed process

$q_i(x_i \mid h)$	Conditional probability of strategy x_i, given that i experiments in the state h
μ^ε	Stationary distribution
$r_{hh'}$	Resistance of the directed edge from h to h'
$\Im(x)$	Minimal resistance tree rooted at h_x
$\rho(x)$	Stochastic potential of h_x
$A_k^j(t)$	Individual k's attraction to strategy x^j in period t
λ	Payoff sensitivity
δ	Degree of hypothetical reinforcement
σ	Discount factor for previous observations
ϕ	Discount factor for previous attractions

Sammanfattning på svenska

Denna avhandling innehåller fyra uppsatser som undersöker vilka beteenden som kan förväntas när populationer av begränsat rationella individer interagerar under lång tid. Dessa individers strategival antas bestämda av enkla beslutsregler som använder information om historiska strategival.

De tre första uppsatserna har en gemensam inramning snarlik den i Young (1993, 1998). I varje period dras en individ från var och en av n populationer för att spela ett ändligt n-personersspel. Varje individ väljer en strategi enligt en särskild beslutsregel efter att ha observerat ett urval utan återläggning från en ändlig historia av tidigare strategival. Med en viss sannolikhet gör individerna dessutom misstag och spelar en slumpmässigt utvald strategi.

I det första kapitlet använder samtliga individer en beslutsregel som innebär imitation av den mest framgångsrika strategin i ett urval av den egna populationens tidigare strategival. I det andra kapitlet använder samtliga individer en beslutsregel som föreskriver spel av en strategi som ger åtminstone lika hög förväntad avkastning som en empirisk fördelning baserad på ett urval av samtliga populationers tidigare strategival. Det tredje kapitlet är en utvidgning som tillåter heterogena populationer, där både det första och det andra kapitlets beslutsregler finns representerade, tillsammans med en tredje beslutsregel som innebär spel av en strategi som maximerar den förväntade avkastningen givet en empirisk fördelning baserad på ett urval av motståndarpopulationernas tidigare strategival.

I samtliga av de tre första kapitlen erhålls generella konvergensresultat för ändliga n-personersspel; enbart strategier i särskilda mängder spelas på lång sikt när sannoliheten för misstag går mot noll. Dessa resultat illustreras därefter i tvåpersoners koordinationsspel.

I den fjärde och sista uppsatsen är inramningen och frågeställningen något annorlunda. I var och en av ett ändligt antal perioder dras samtliga individer från stor population med ett ändligt antal individer och indelas slumpmässigt i par för att spela ett tvåpersonersspel. Individerna använder en beslutsregel från en stor klass av regler som kan representeras av Camerers och Hos (1999) modell. Med hjälp av Monte Carlo-simuleringar undersöks huruvida det finns en beslutsregel som är *evolutionärt stabil* i den mening att om den används av en tillräckligt stor andel av populationen, så ger den en högre förväntad avkastning än varje annan regel som används av resten av populationen. Resultaten från fyra olika spel indikerar att enbart den beslutsregel som tar full hänsyn även till den hypotetiska avkastningen från strategier som inte väljs är evolutionärt stabil i nästan samtliga fall.

EFI
The Economic Research Institute

Reports since 1995
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2001

Andersson, P., Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.
Björklund, C., Work Motivation - Studies of its Determinants and Outcomes.
Center for Management and Organization 50 (1951-2001).
Charpentier, C., Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämnds-reformen.
Dahlén, M., Marketing on the Web - Empirical Studies of Advertising and Promotion Effectiveness.
Ekelund, M., Competition and Innovation in the Swedish Pharmaceutical Market.
Ericsson, D., Kreativitetsmysteriet – Ledtrådar till arbetslivets kreativisering och skrivandets metafysik.
Eriksson, R., Price Responses to Changes in Costs and Demand.
Frisell, L., Information and Politics.
Giordani, P., Essays in Monetary Economics and Applied Econometrics.
Gustavsson, P., Essays on Trade, Growth and Applied Econometrics.
Hedlund, A., Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.
Hill, M., Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.
Hvenmark, J.,Varför slocknar elden? Om utbrändhet bland chefer i ideella organisationer.
Hägglund, P.B., Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.
Höök, P., Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.
Johansson, C., Styrning för samordning.
Kjellberg, H., Organising Distribution - Hakonbolaget and the efforts to rationalise food distribution, 1940-1960.
Lange, F. och Wahlund, R., Category Management – När konsumenten är manager.
Liljenberg, A., Customer-geared competition – A socio-Austrian explanation of Tertius Gaudens.
Lindkvist, B., Kunskapsöverföring mellan produktutvecklingsprojekt.
Ljunggren, U., Nyckeltal i grundskolan i Stockholms stad före och efter stadsdels-nämndsreformen.
Läkemedel – Kostnad eller resurs för sjukvården? (red. Jönsson, B.,).
Löf, M., On Seasonality and Cointegration.
Martensen, K., Essays on Entry Externalities and Market Segmentation.
Matros, A., Stochastic Stability and Equilibrium Selection in Games.
Mårtensson, P., Management Processes – An Information Perspective on Managerial Work.
Nilsson, A., Market Transparency.
Persson, B., Essays on Altruism and Health Care Markets.

Skoglund, J., Essays on Random Effects Models and GARCH.
Strand, N., Empirical Studies of Pricing.
Thorén, B., Stadsdelsnämndsreformen och det ekonomiska styrsystemet - Om budget-avvikelser.

2000

Berg-Suurwee, U., Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.
Bergkvist, L., Advertising Effectiveness Measurement: Intermediate Constructs and Measures.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Koncernchefen och ägarna.
Bornefalk, A., Essays on Social Conflict and Reform.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform.
Edman, J., Information Use and Decision Making in Groups.
Emling, E., Svenskt familjeföretagande.
Ericson, M., Strategi, kalkyl, känsla.
Gunnarsson, J., Wahlund, R., Flink, H., Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.
Hellman, N., Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.
Hyll, M., Essays on the Term Structure of Interest Rates.
Håkansson, P., Beyond Private Label – The Strategic View on Distributor Own Brands.
I huvudet på kunden. Söderlund, M., (red).
Karlsson Stider, A., Familjen och firman.
Ljunggren, U., Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.
Ludvigsen, J., The International Networking between European Logistical Operators.
Nittmar, H., Produktutveckling i samarbete – Strukturförändring vid införande av nya Informationssystem.
Robertsson, G., International Portfolio Choice and Trading Behavior.
Schwarz, B., Weinberg, S., Serviceproduktion och kostnader – att söka orsaker till kommunala skillnader.
Stenström, E., Konstiga företag.
Styrning av team och processer – Teoretiska perspektiv och fallstudier. Bengtsson, L., Lind, J., Samuelson, L.A., (red).
Sweet, S., Industrial Change Towards Environmental Sustainability - The Case of Replacing Chloroflouorocarbons.
Tamm Hallström, K., Kampen för auktoritet - Standardiseringsorganisationer i arbete.

1999

Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Experto Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Flink, H., Gunnarsson, J., Wahlund, R., Svenska hushållens sparande och skuldsättning– ett konsumentbeteende-perspektiv.
Gunnarsson, J., Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
Hamrefors, S., Spontaneous Environmental Scanning.

Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Japanese Production Management in Sunrise or Sunset. Karlsson, C., (red).
Jönsson, B., Jönsson, L., Kobelt, G., Modelling Disease Progression And the Effect of Treatment in Secondary Progressive MS. Research Report.
Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., Indikatorer i grundskolan i Stockholms stad före stadsdels-nämndsreformen – en kartläggning.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Mägi, A., Store Loyalty? An Empirical Study of Grocery Shopping.
Mölleryd, B.G., Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
Nilsson, K., Ledtider för ledningsinformation.
Osynlig Företagsledning. Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red).
Rognes, J., Telecommuting – Organisational Impact of Home Based – Telecommuting.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Spagnolo, G., Essays on Managerial Incentives and Product-Market Competition.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Svedberg Nilsson, K., Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
Söderström, U., Monetary Policy under Uncertainty.
Werr, A., The Language of Change The Roles of Methods in the Work of Management Consultants.
Wijkström, F., Svenskt organisationsliv – Framväxten av en ideell sektor.

1998

Andersson, M., On Testing and Forecasting in Fractionally Integrated Time Series Models.
Berg-Suurwee, U., Styrning av kultur- och fritidsförvaltning innan stadsdelsnämnds-reformen.
Berg-Suurwee, U., Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdels-nämndsreformen.
Bergström, F., Essays on the Political Economy of Industrial Policy.
Bild, M., Valuation of Takeovers.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.
Eriksson-Skoog, G., The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.
Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.
Ioannidis, D., I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.
Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.
Levin, J., Essays in Company Valuation.
Ljunggren, U., Styrning av grundskolan i Stockholms stad innan stadsdelsnämnds-reformen.
Mattsson, S., Från stat till marknad – effekter på nätverksrelationer vid en bolagise-ringsreform.
Nyberg, A., Innovation in Distribution Channels – An Evolutionary Approach.

Olsson, P., Studies in Company Valuation.
Reneby, J., Pricing Corporate Securities.
Roszbach, K., Essays on Banking Credit and Interest Rates.
Runsten, M., The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.
Segendorff, B., Essays on Bargaining and Delegation.
Sjöberg, L., Bagozzi, R., Ingvar, D.H., Will and Economic Behavior.
Sjögren, A., Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.
Studier i kostnadsintäktsanalys. Jennergren, P., (red)
Söderholm, J., Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.
Thorburn, K., Cash Auction Bankruptcy and Corporate Restructuring
Wijkström, F., Different Faces of Civil Society.
Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.

1997

Alexius, A., Essays on Exchange Rates, Prices and Interest Rates.
Andersson, B., Essays on the Swedish Electricity Market.
Berggren, N., Essays in Constitutional Economics.
Ericsson, J., Credit Risk in Corporate Securities and Derivatives. Valuation and Optimal Capital Structure Choice.
Charpentier, C., Budgeteringens roller, aktörer och effekter. En studie av budgetprocesserna i en offentlig organisation.
De Geer, H., Silfverberg, G., Citizens'Trust and Authorities' Choices, A Report from The Fourth International Conference on Ethics in the Public Service, Stockholm June 15-18, 1994.
Friberg, R., Prices, Profits and Exchange Rates.
Från optionsprissättning till konkurslagstiftning. Bergström, C., Björk, T., (red)
Hagerud, G.E., A New Non-Linear GARCH Model.
Haksar, A., Environmental Effects of Economywide Policies: Case Studies of Costa Rica and Sri Lanka.
He, C., Statistical Properties of Garch Processes.
Holmgren, M., Datorbaserat kontrollrum inom processindustrin; erfarenheter i ett tidsperspektiv.
Jennergren, P., Tutorial on the McKinsey Model for Valuation of Companies.
Lagerlöf, J., Essays on Political Economy, Information, and Welfare.
Lange, F., Wahlund, R., Planerade och oplanerade köp - Konsumenternas planering och köp av dagligvaror.
Löthgren, M., Essays on Efficiency and Productivity; Contributions on Bootstrap, DEA and Stochastic Frontier Models.
Nilsson, H.E., Using Conceptual Data Modelling in Management Accounting: A Case Study.
Sjöberg, L., Ramsberg, J., En analys av en samhällsekonomisk bedömning av ändrade säkerhetsföreskrifter rörande heta arbeten.
Säfvenblad, P., Price Formation in Multi-Asset Securities Markets.
Sällström, S., On the Dynamics of Price Quality.
Södergren, B., På väg mot en horisontell organisation? Erfarenheter från näringslivet av decentralisering och därefter.
Tambour, M., Essays on Performance Measurement in Health Care.
Thorén, B., Berg-Surwee, U., Områdesarbete i Östra Hökarängen - ett försök att studera effekter av decentralisering.

Zhang Gang., Chinese Rural Enterprises Between Plan and Market.
Åhlström, P., Sequences in the Profess of Adopting Lean Production.
Åkesson, G., Företagsledning i strategiskt vakuum. Om aktörer och förändrings-processer.
Åsbrink, S., Nonlinearities and Regime Shifts in Financial Time Series.

1996

Advancing your Business. People and Information Systems in Concert. Lundeberg, M., Sundgren, B (red).
Att föra verksamheten framåt. Människor och informationssystem i samverkan. red. Lundeberg, M., Sundgren, B.
Andersson, P., Concurrence, Transition and Evolution - Perspectives of Industrial Marketing Change Processes.
Andersson, P., The Emergence and Change of Pharmacia Biotech 1959-1995. The Power of the Slow Flow and the Drama of Great Events.
Asplund, M., Essays in Industrial Economics.
Delmar, F., Entrepreneurial Behavior & Business Performance.
Edlund, L., The Marriage Market: How Do You Compare?
Gunnarsson, J., Three Studies of Financial Behavior. Research Report.
Hedborg, A., Studies of Framing, Judgment and Choice.
Holgersson, C., Höök, P., Ledarutveckling för kvinnor - Uppföljning av en satsning på Volvo - Research Report.
Holmberg, C., Stores and Consumers - Two Perspectives on Food Purchasing.
Håkansson, P., Wahlund, R., Varumärken. Från teori till praktik.
Karlsson, A., The Family Business as an Heirloom. Research Report.
Linghag, S., Man är handelsstudent. Research Report.
Molin, J., Essays on Corporate Finance and Governance.
Mägi, A., The French Food Retailing Industry - A Descriptive Study.
Mölleryd, B., Så byggdes en världsindustri - Entreprenörskapets betydelse för svensk mobiltelefoni. Research Report.
Nielsen, S., Omkostningskalkulation for avancerede produktions-omgivelser - en sammenligning af stokastiske og deterministiske omkost-ningskalkulationsmodeller.
Normark, P., Danielsson, L., Larsson, A., Lundblad, P., Kooperativa nyckeltal. Research Report.
Sandin, R., Heterogeneity in Oligopoly: Theories and Tests.
Sandén-Håkansson, U., Från kampanjmål till mediemix - en studie av samarbete mellan annonsörer, reklambyråer och mediebyråer. Research Report.
Stein, J., Söderlund, M., Framgång i arbetet, strategier för att utföra arbetsuppgifterna, arbetsuppgifternas karaktär och utbildningskvalitet. En empirisk studie av civil-ekonomer. Research Report.
Strömberg, P., Thorburn, K., An Empirical Investigation of Swedish Corporations in Liquidation Bankruptcy. Research Report.
Söderlund, M., Och ge oss den nöjda kunden. En studie av kundtillfredsställelse och dess orsaker och effekter. Research Report.
Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1995. Research Report.
Ulfsdotter, U., Internationalisering för expansion eller hemmamarknadsförsvar? De nordiska marknaderna för fruktyoghurt 1982-1994.
Westelius, A., A Study of Patterns of Communication in Management Accounting and Control Projects.
Wijkström, F., Den svenska ideella sektorn och pengarna. Research Report.
Örtendahl, M., Health and Time - A Problem of Discounting.