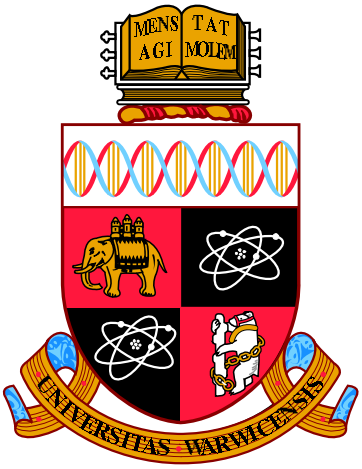


Optimal Policy for Large Herding Population

by

Gian Lorenzo Spisso

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Centre for Complexity Science

September 2019

THE UNIVERSITY OF
WARWICK

Contents

Acknowledgments

Declarations

This dissertation is the result of my own work and, without exception, includes nothing which is the outcome of collaboration.

Abstract

This thesis proposes a reduced-form model of herding for a large population of boundedly rational agents receiving logistic preference shocks. Agents are called at random times to revise a binary choice — smoke or not, vote for or against a party, buy or sell an asset — as they try to align to other agents choices and the policy set by an external planner. I show that the resulting Markov stationary distribution places most probability on configurations of agents choices that maximize the sum of individual utilities, similar to Kandori et al. (2008). Reducing the dimensionality of the economy — by aggregating configurations based on the average value of agents choice — shows that, in fact, since less likely configuration are more frequent, agents spend a long time away from the most likely ones. In particular, when agents are rational and care about coordination, two *long-lived equilibria* emerge. I study the problem of a planner who has an exogenous preference for one choice and sets a policy to minimize discounted expected costs. An algorithm by Ross (1983) is applied to compute the optimal policy under different assumption to show that the optimal policy is, in general, non-monotonic and at times discontinuous, with a region where the planner "gives up" and sets a policy close to zero. It is shown that the planner maximum policy is always exerted when at least half of the population is opposed to the planner preferred choice. Further, I give bounds on the marginal costs of policy that determines whether long-lived equilibria are still present under the optimal policy.

Chapter 1

Introduction

This thesis deals with the dynamic behavior of a large population of agents attempting to coordinate on two alternative choices — buy or sell, smoke or don't, vote for one party or another, adhering or not to a convention or social norm — displaying a multiplicity of equilibria. An external agent, a government or a planner, enacts a policy — by raising a tax, changing the current interest rate or by running an advertisment campaign — to influence the population toward one of the available equilibria.

Coordination games of this sort underlie many situation of economic interest such as: herding in financial markets (Devenow and Welch, 1996), growth or poverty traps (Krugman, 1998), elections (Feddersen and Pesendorfer, 1997), opinion formation and social norms (Young, 1993) , investment hold-up and capital gathering (Akerlof et al., 2018, 2019) , expectation formations (Golub and Morris, 2017). At the core is the fact that people would benefit from being able to coordinate on an outcome but lack a mechanism to coordinate or are exposed to unmediated externalieties, often leading to multiple equilibria. As consequence, game theory work has focused either on providing *equilibrium refinements*, progressive tightening of equilibria requirement toward more sensible ones, and *equilibrium selection*, the description of how one equilibrium is selected either dynamically, through repetition or learning, or in the limit of some relevant parameter. During the ninties and early two-thousands, the latter topic has been explored by evolutionary modeling. Borrowing tools from system biology and statistical physics game theorists have proposed different notions of evolutionary equilibrium as an attempt to explain how one equilibrium is selected among many.

The present work, whilst sharing many features of the evolutionary game theory literature, tackles multiple equilibria from a different point of view, namely

how transition across equilibria happen and *how to optimally drive agents* across them. Out of the three typical assumptions of evolutionary game theory, two are maintained here in the same spirit: first, *inertia* — agents in our model don't react instantaneously to their environment, but rather with a delay. This lag is motivated by the fact that agents live in a complicated world and have a limited amount of attention to give to a certain choice, in turn facing opportunity cost whenever they have to research and decide which strategy or choice to make. The usual device (see for example Blume (1993)) — an i.i.d. exponential clock with mean equal to the advantage of revising the choice — determines when agents get to act . The time passing between two revisions will depend on how convenient it is to switch: when the gain from switching is negliglible agents will be less likely to activate and re-evaluate their current choice; conversely, the higher the gain from switching the sooner agent will act.

The second assumption is *myopia* — agents do not forecast the behavior of other agents in the future nor how they might react to their choices. When agents make a choice they only consider the current choice of others and the planner policy at any given time. This is motivated by cognitive constraints that agents have once they have decided to review their choice, hence agents do not take into account the long term implication of their choices nor they attempt at forecasting[1] the long term behavior of others and how other will react to their choices and so on. In practice, this means that agents only maximise their current utility.

The third typical assumption *mutations* — agents put non zero probability on every possible strategy — is introduced here with a twist: agents receive a random shock to their utility *every time they are called to make choice*. Continual arrival of individual idiosyncratic shocks induces a Markovian dynamic allowing the study of the population behavior as a stochastic process. Shocks such as this have been motivated in a variety of way. One way is to think that agents make mistakes, leading them to act against the deterministic component of their utility. Another way is to think that the population changes over time: whenever a new choice is made one player leaves and is replaced by a new one with different preferences. Otherwise, agents could be seen as experimenting, again agents with limited cognitive ability proxy for more complicated strategy by randomly acting in a different way that would otherwise be predicted by their utility. One more interpretation, is that shocks are unobservable factors that enter the utility function[2]. Finally, agents preference might fluctuate and be changing exogenously over time. The latter will

[1] See conclusion for a discussion of the effect of iterated expectations

[2] See later for a more detailed explanation in the Random Utility Model section of the preliminaries.

be the preferred explanation of shocks.

These three hypothesis come together to describe dynamic, *boundedly rational agents*. Bounded rationality is crucial, both because of its intuitive and practical appeal. Intuitively, because we rarely see agents acting in the fully rational way of classical economics modeling. Practically, because without some sort of delay and some exogenous motivation to change strategy we would be bound to a static world where players immediatly reach the equilibrium, leaving us with the issue of how an equilibrium is selected. The way in which we model the population allows all equilibria to be considered and to study the dynamic of how transition happens across them. Notice, that while agents do not forecast the future and therefore do not immediatly reach equilibrium, their average behavior will be coherent with utility maximization since shocks are i.i.d. with zero mean. Further, in the limit of small shocks the fully rational behavior can be recovered.

The last element needed to describe agents behavior is their utility function, which encapsulates two elements: a desire to conform to the average choice of all other agents and to align with the planner policy. The latter has an easy interpretation: the planner can impose a monetary tax or non-monetary cost — such as graphic pictures on cigarette packs — to deter people from making a certain choice and at the same time incentivize agents to make the opposite decision. The former is another application of bounded rationality and can be motivated in a few ways. One can see the desire to align with the average population choice as an exogenously given preference for imitation. It should be noted though, that preference for imitation might arise from rational concerns. For example, concern for relative position with concave utility (Clark and Oswald, 1998) can lead to imitation. Some herding mechanism, see (Devenow and Welch, 1996) for a review, also result in imitation of other choice and discard of private information. Lastly, agents might simply benefit directly from "jumping on the bandwagon", for example when agents choose between two party and having voted for the final winner might result in a reward. We shall abstract from the fine details and assume agents have a preference for aligning with the average position in the population, thinking of this model as reduced-form model of herding.

Chapter 3 describes the microfoundations and global behavior of the population under some fixed policy from the planner. The mathematical formalism used is that of discrete time, discrete state space, Markov chains. The object of interest are the transition probabilities dictated by agents utilities and the long run behavior of the population described by the stationary distribution of the chain. A dimensionality reduction technique, *lumping*, is used to naturally connect the popu-

lation behavior with the behavior of the average. The first result is to show that the configurations that have higher probability are those that maximise the utilitarian welfare, that is, the sum of all utilities. It is not surprising that this is the case, Kandori et al. (2008) have shown the same result in a slightly different setup. But we also show, through lumping, that their prediction about the most likely states being the one where the population spends the most time is not always true, due to the fact that less likely configuration might outnumber more likely ones. Secondly, it is shown that subsets of population choice act as *basins of attraction*, which we term *long lived equilibria*. Once the population is within this set they spend a disproporionate amount of time within it, which is exponential in the size of the population. The mathematical setup follows closely Bovier and den Hollander (2015), with the crucial difference of how transition probabilities are derived from the behavior of the single agents which is a boundedly rational version of the model presented in Durlauf (1996).

The following chapters deal with the optimization problem faced by the planner. The planner has an exogenous preference for one of the two choices that agents can make and faces quadratic costs proportional to the distance of the average population choice from its preference. The planner can choose a continuous policy which affects all agents rewarding agents that align with it and punishing those that are not. The intensity of the policy is the choice variable for the planner and symmetrically determines how much player are rewarded or punished. For every unit of policy the planner pays an associated quadratic cost. Over an infinite time horizon the planner would like to pick the policy that minimize its expected discounted costs. The tool used are those of stochastic dynamic programming, presenting two applications of the Policy Improvement Algorithm discussed in Ross (1983), allowing for the numerical computation of the optimal policy. One can motivate this type of planner as a benevolent planner that wishes to cull smoking behavior in the population. Due to the peer pressure that smokers exercise on each other, when a majority of people smoke it might be very hard to convince them to do otherwise. Yet again, the planner might be an incumbent in an election and wishes to retain the curren majority of people voting for him. The optimal policy presented here suit both type of situations: changing the status quo or maintaining it.

The planner problem is solved under two different assumptions. In Chapter 4 it is assumed that the planner can only observe the average position of the agents every so often. This exogenously given lag reflects the fact that sampling the population might be costly. Under this assumption the planner observes the population behaving according to the stationary distribution. A change in the policy

alters the stationary distribution shifting probability toward the planner preferred state. I show that the unique optimal policy value can be seen as an average of the costs faced by the planner weighted by the likelihood of the population reaching any given configuration. Further, I show that the first order condition for the planner problem relates the optimal policy to the skewness and variance of the stationary distribution, which are affected by the parameters that govern agents behavior. I numerically show that the optimal policy may or may not be monotonic in the variance of shocks depending on how much value is placed by agents on coordination. The other important question answered is whether long lived equilibria survive after the optimal policy is introduced. The answer depends on the marginal costs of applying the policy: high marginal costs entail that the attracting regions survive, whereas for sufficiently low cost the largest possible amount of policy is optimal and this is enough to make long lived equilibria vanish. Bounds are provided for both cases.

In Chapter 5 it is assumed that the planner can monitor the population average position at every time step. As a result, optimal policy is now state dependent: for every possible configuration of choices in the population the planner has an optimal response. I provide a proposition relating the optimal policy to its first order condition which depends on the transition probabilities. In general, the optimal policy is not monotonic. The planner wishes to raise the intensity of the policy as the number of people aligning with its goal decreases, but only up to a certain point. Past it, the planner de-escalates the amount of policy. This happens because the more agents expouse the opposite position the stronger their critical mass pulls the population into one of the long lived equilibria. Past some critical threshold the gain from additional unit of policy is smaller then the cost of implementing it. Further, the reduction need not to be "smooth", in some cases the planner wishes to suddenly and drastically reduce its policy, effectively "giving up" once the majority becomes too large.

There are roughly three possible regimes for the optimal policy. In population with low desire to coordinate and large preference shocks the policy will be monotonic, weaker when people align with the planner and stronger when people aren't. This is intuitively explained by the fact that under no policy the stationary distribution for the population presents a unique long-lived equilibria and stronger policy shifts this basin closer to the planner preferred state. In populations where the coordination motive is strong and shocks are large, the policy becomes non-monotonic, first rising and then falling in absolute value as the population average position moves away from the planner target. When shocks become small — meaning that

agents pick more often the rational choice and therefore become more "sticky" when they are already aligned with the majority — the policy is non-monotonic and may exhibit a "give up" zone for the planner where the policy suddenly drops close to zero. How quick this jump is depend on the population size: for larger population the jump will be small followed by a steep decrease of the policy; for smaller population the jump will be larger. Finally, which of these three regimes manifest depends on the size of marginal costs: low marginal cost cause policy to be larger and progressively more monotonic, whereas large costs lead to smaller and markedly non-monotonic ones. The main result I prove, for a sufficiently large population, is that the largest amount of policy is provided when more than half of the agents have taken position opposing the planner. Again, whether long-lived equilibria survives the optimal policy is explored numerically and an analytical formula for the new stationary distribution under the state dependent optimal policy is provided.

1.1 Summary of main results

Here I summarize the main contribution of the thesis. In Chapter 3 I provide the first description of a system of boundedly rational agents making dichotomic choices and receiving logistic preference shocks as they play an infinite horizon coordination game. An overseeing agent — the planner — can influence the coordination by introducing a policy. This model expands on many previous work in evolutionary game theory, introducing a new way of interpreting the presence of multiple equilibria as attracting region of the state space of a Markov chain. These attracting regions can be thought as "equilibria" in the sense that the system spends a long time in them. In the chapter I characterize the long-run behavior of all agents as the stationary distribution of the chain placing most probability on those configurations — the vector of all agents choices — that maximise total utility. I also show how the dimensionality of such a game can be reduced through *lumping* techniques. Proposition 1 and 2 give the analytical formula for the stationary distribution of the population configuration and the lumped stationary distribution of the average choice in the population. Proposition 3 defines for what value of a fixed policy the lumped stationary distribution is bi-modal. The region depends on population parameters describing the strength of the coordination motive and the variability of the shock. Propositon 4 shows that the minima of the potential function identify the region of long-lived equilibria and Proposition 5 describes the time needed to escape and enter into the attracting region of a long-lived equilibria .

I then proceed to build and solve the problem of an external planner who

whishes to drive the population towards one of its preferred equilibria using the formalism of stochastic optimization. The resulting policy is characterized both analytically and numerically to show the presence of distinct monotonic and non-monotonic regimes that depend on both the marginal costs of the planner and on the two parameters that describe agents behavior: strength of the coordination motives and variability of the preference shocks. I provide a proof using coupling techniques to show that the value function is increasing over the state space.

In Chapter 4 the planner problem is setup as a stochastic optimization problem under the assumption that the planner only observe agents with a lag, meaning that the behavior of the population is dictated by the lumped stationary distribution. The object of interest is the value function of the planner, associating every element of the lumped state space with the minimized discounted expected costs. Under the maintained assumption Proposition 6 shows that the value function is convex in the policy. The trick is to use the representation the value function as an average over all states weighted by the stationary distribution, which takes the form of a geometric series. Since there's a one to one relationship between the value function and the optimal policy, we can relate the minima of the former to the optimal values. Proposition 7 gives the first order condition that the optimal policy must satisfy and shows that it depends on the skewness and variance of the lumped stationary distribution. It is also shown numerically that the optimal policy as a function of the variability of agents preference shocks has a monotonic and non-monotonic regime that depend on the strength of the coordination motive. Lastly, Proposition 8 answer the question of whether long-lived equilibria survive with the application of the optimal policy, by giving upper and lower bounds on the policy marginal costs that guarantee survival or deletion.

In Chapter 5 the planner problem is solved under the assumption that the average choice of all agents can be observed at any time. The value function is no longer convex in the policy, although it remains to be shown whether this can be shown analytically, and the optimal policy now becomes state dependent. Proposition 9 gives the first order condition relating the optimal policy to the transition probabilities of the lumped chain and the variation in the value function. Optimal policy is in general non-monotonic over the lumped state space and Proposition 11 gives bounds on costs guaranteeing either regime. A key mathematical result that is needed is that the value function is increasing over the state space. Proposition 12 proves that this is the case, providing a proof by coupling that also gives useful upper and lower bounds on the variation of the value function. Lastly, Proposition 13 shows that, for sufficiently large population, the peak of the optimal policy hap-

pens after at least half of the agents take the opposite position the planner prefers. This is shown numerically to be always the case, even for small populations.

1.2 Literature Review

The work in this thesis stems from the meeting point of three different literatures: the economic theories on imitation and herding, the evolutionary game theory literature and the application of statistical mechanics to socio-economic phenomenon. Here I try to review the most important paper in these three literatures, highlighting similitudes and differences with my work. The mathematical tools that bring all of these together are given by the Markov chain literature and in particular its application to interacting particle systems, which are described in Section 2.2.

1.2.1 Interactions, relative standing and herding

Economics and sociology have often dealt with the topic of how agents' action affect each other. One might say that this is in fact the main focus of both disciplines. Here we focus on a specific strand of literature, which attempts to mathematically derive the *aggregate behavior* of the population starting from the rule governing the individual agents.

The first example of such models is due to Schelling's study of segregation. Schelling notes that segregation may be due to social structures : organizational practices, specialized communications or, more generally, correlation with a non-random variable. Sometimes, though, segregation might arise from the interplay of individual choices.

Schelling (1971) studies segregation by presenting a simulation of agents from distinct population attempting to relocate themselves spatially based on a preference for their surroundings. An analytical model of the same phenomenon with discretised space[3] is also discussed. The general message is that, even in the presence of simple rules, the interplay of individual choices produces a vast array of phenomena such as separation, patterning, density, vacancy and the appearance of drastic "tipping points". Further, the paper argues that there is no simple correspondence from individual incentives to aggregate behavior. Conversely, inference of individual motives cannot be drawn from the aggregate patterns.

Schelling (1973) deals with the study of binary choices with externalities. The example used is that of the adoption of helmets by hockey players at a time

[3]This is effectively a conserved order parameter two-dimensional Ising model on the square lattice, a model which originates in the field of statistical mechanics.

when hockey rules did not mandate the use of one. The convention of not using helmets, Schelling convincigly argues, is uphold by the effect of peer pressure: since no player is using them — despite it being clearly very beneficial — no one is willing to deviate from the accepted standard, perhaps for fear of ridicule or of gaining a disadvantage. Again, the message of the paper is that from different hypothesis on the fine details of how the interaction among agents happen, a large variety of situations might emerge.

What schelling describes in his 1973 paper is what economics literature will call — especially in the context of financial markets — *herding*: a situation where agents behave in a certain way — sometimes against their own interest — due to other agents already taking the same action.

A good review of the rational herding literature — models in which agents optimal choice is to herd — is Devenow and Welch (1996), that gives the following definition:

> "In its most general form, herding could be defined as behavior patterns that are correlated across individuals. But, if many investors are purchasing 'hot' stocks, it could just be due to correlated information arrival in independently acting investors. The notion of 'herding' we consider instead is one which can lead to systematic erroneous (i.e., sub-optimal relative to the best aggregate choice) decision-making by entire populations. In this sense, herding is closely linked to such distinct phenomena as imperfect expectations, fickle changes without much new information, bubbles, fads, frenzies, and sun-spot equilibria."

Models in the rational herding literature are usually based on either payoff externalities; principal-agent models showing that managers might have an incentive to 'hide in the herd'; or, informational cascades model where action from prior agents leads others to ignore private information and imitate others.

Yet another important mechanism that leads to correlated behavior is the role of relative position, explored by Cole et al. (1992). In their model 'status' is treated as a ranking device that determines how people fare in the non-market sector of the economy. They show that the existence of a non-market sector might endogenously generate a concern for relative position in income distribution, leading to higher income implying higher status.

On more general line, the work of Clark and Oswald (1998) assumes that relative standing is a concern and shows that when agents have concave utility this is sufficient to ingenerate herding, with people acting against their own best interest

in an attempt to "keep up with the Joneses". The logic driving this result is that, as other agents alter their status, marginal utility of status also changes.

The last paper that deserves to be mentioned is a recent work by Golub and Morris (2017). In their paper they examine the development of higher-order expectations for agents placed on a general network. Agents are in fact attempting to coordinate on beliefs as they receive a random signal on the state of the world, using a utility function that shares the same type of quadratic loss in the distance from the position of nearby agents as the one used in this thesis. Their main result is to show — using Markov methods — how even a small amount of optimism about other agents' expectations leads to a 'contagion of optimism'. Conversely, the lack of optimism leads to a 'tyranny of the least-informed' as agents end up coordinating on the prior expectations of the agent with worst private information.

1.2.2 Evolutionary Game Theory

An evolutionary model formalizes the process of learning a Nash equilibrium by a large population of myopic and unsophisticated agents. The central notion justifying the term "evolutionary" is the concept of an *Evolutionary Stable Strategy* (ESS) first introduced by Smith and Price (1973). The idea is that a stable pattern in a population is evolutionary stable if it cannot be invaded by a "mutant" pattern. Any ESS is a Nash equilibrium, but the converse is not necessarily true. Despite representing a useful restriction of a Nash equilibrium, ESS might fail to exists. An overview of the early research surrounding this concept can be found in Mailath and J. (1992). Fundenberg and Maskin (1990) shows that evolutionary stability imply a notion of efficiency as long as players make sufficiently small mistakes: this new notion is known as "Stochastically Stable Strategy". The work on conventions[4] by Young (1993) shows that the for a particular class of 2x2 games the equilibrium that are stochastically stable are equivalent to the risk-dominant equilibrium.

The work in Kandori et al. (1993) introduces the notion of repeated shocks at the population level leading to what they term "long-run equilibria". In 2x2 symmetric games with two symmetric Nash equilibria these match the same concept of risk-dominant equilibrium of Harsanyi and Selten (1988) and in particular, with equal level of security, the Pareto dominant Nash equilibrium is selected. The evolutionary process presented in their previous work is extended in Kandori and Rob (1995) to show that long-run equilibria can be computed algorithmically for $n \times n$ games and that this is unique even when multiple static equilibria exist.

[4] An overall review of the work on social norm and conventions is Burke et al. (2011).

The work of Ellison (1993, 2000) is instead focused on the amount of time that these models take to attain equilibrium through experimentation and mistakes, showing that when interactions happen at long-range the time to attain the unique long-run equilibrium might be long and dictated by the history and the initial conditions. When interactions are short-range instead equilibrium is attained much faster and the history of the players adjustment is less relevant[5].

Friedman (1998) discusses possible applications of evolutionary game theoretic model and argument how the usual restriction on agents rationality might be less strong than usually thought.

Kandori et al. (2008) studies a decentralized trading process where agents are affected by persistent random shocks due to agents' random utility maximization. In this model agents group meet randomly and exchange indivisible durable goods among each other. Once they meet a new allocation of the goods of the group is picked at random and agents accept it if the new allocation provides a higher utility to all of them. The main result in this paper is show that the stationary distribution for myopic agents receiving logistic shocks has an exponential form proportional to the sum of agents utilities[6]. Compared to this work this thesis features shocks that are identically distributed, rather than have a possibly different variability for each agent and abstracts from matching probabilities. The paper is extended here by the introduction of an external planner incentive. Another difference, is that the work by Kandori, Serrano and Volij fails to notice that while most probability is placed on highest utilitarian welfare configuration, these are only a few out of the many possible configurations and these configurations might be outnumbered by others with lower welfare: reducing the dimensionality of the system through lumping allows us to figure out which of these two effects — higher utility or higher frequency of a configuration — wins.

1.2.3 Statistical Mechanics applications

Economics has a long tradition of retooling physical models and techniques. Statistical Mechanics is the field that deals with the behavior of large ensamble of particles whose dynamics is treated as stochastic. Stochasticity is seen there as a simplification of underlying deterministic dynamics that might be too complex to keep track

[5]This is well known in the statistical mechanics literature, and is referred to as the fact that so called mean-field models, where all particles interact with all others, present metastability, which disappears once interaction become short range (Bouchaud, 2013).

[6]This extends Volij et al. (2004) by showing that the concept of minimum envy allocation can be replaced by the general principle that evolutionary dynamics with logit noise maximise the aggregate utility level.

off. One of the biggest hurdle in transposing its formalization is that the behavior of physical systems obeys aggregate laws that can be derived experimentally and imposed as constraints on the aggregate behavior of single particles. This is rarely the case in economics, hence research in this direction has been devoted to motivating stochasticity at the individual level. Yet another challenge is posed by the fact that, since the exposition of the so called Lucas (1976) critique, economists strove to include expectations of agents as part of the construction of sensible *microfoundations*. But physical particles are usually treated as reacting only to their current environment and not forecasting the future. This is where the statistical mechanics applications merged with evolutionary game theory assumptions that agent might react myopically to their environment and rational behavior might emerge at the aggregate level as a consequence of the underlying incentives.

In particular, the economics literature I will discuss here has borrowed heavily from the formalism of the Ising model for ferromagnetism – describing how magnetic dipole moments on a lattice align to each other to generate a magnetic field – and its mean field counterpart the Curie-Weiss model, which are some of the first model used to describe and study phase transition in physical matter. A full review of the Curie-Weiss model can be found in Kochmański et al. (2013), while an introduction to the Ising model is in Baxter (2016).

The main paper that applies the logic of statistical mechanics to game theory is the seminal work of Blume (1993), *The Statistical Mechanics of Strategic Interaction*. Here, player interact with each other on a lattice, meaning that each player only plays a game with a finite set of neighbours, but every two players indirectly interact with all others through chain of direct interactions. The paper describes how such construction works under different stochastic strategy revision processes. It also details the stationary distribution and limiting behavior of the underlying Markov chain, in particular for coordination games. The two revision processes considered are *perturbed best response* and *log-linear model*. The former consists of agents that place some positive probability on each outcome when selecting a strategy; the latter is a variation of that specifying the functional form of the log-odds of choice revision. Log-linear model have the exact same implication as the logistic shocks agents receive in this thesis: probability of revising a choice is proportional to the difference in utilities between the outcomes. Blume also introduces the device, common in Interacting Particle Systems literature (see Liggett (1985)) that defines the activation times of agents: to each player is attached an i.i.d. Poisson "alarm clock" which rings at random times. When it does, the agent is activated and reacts to its neighbours current configuration, giving rise to Markovian dynamics. The

main results in Blume's paper show that the log-linear rule gives rise to stationary distributions that have Gibbs form as well as showing that in the limit of small shocks a unique equilibrium is selected and it coincides with the Nash risk-dominant equilibrium when one exists.

Durlauf (1996) work — titled *Statistical Mechanics Approaches to Socio-Economic Behavior* — provides a framework for models of interaction borrowing the formalism of statistical mechanics. Agents maximise a random utility function by making dichotomic choices. The utility function contains a private deterministic component and a social component that includes expectations of all other agents behavior. The social components takes the form of a squared loss function with respect to other agents choices, which can be reduced to the expected average of the population when interactions are global. When the private component takes a linear form it is shown that any linear component that doesn't depend on the agent choice is irrelevant. The main result shows that in the limit of infinite agents with rational expectations, the equilibrium distribution of the model is given by a Gibbs distribution. This result hinges on a fundamental theorem in statistical mechanics Spitzer (1971); Averintsev (1970) showing that these type of stochastic interactions model will in general show the Gibbs distribution. The paper shows that there might be a multiplicity of equilibria of these type of models depending on the product of the parameters for the coordination motive and variability of shocks. This result is extended in Proposition 3 in the current thesis. The paper provides a number of generalizations and applications of the model showing that it can be adapted to encompass endogenous preferences, represent growth and economic development as well as representing games with strategic complementarities. The model used in this thesis is based on the model presented in this paper. The main difference is that the assumption of rational expectation is completely abandoned and the fully forward looking agents are replaced with myopic ones. This has the consequence that the model is a fully dynamic one instead of an equilibrium one and can be studied for a finite number of agents.

Brock and Durlauf (2001) extends the previous paper by introducing a social planner who attempts to maximise the sum of the deterministic utilities of the population and who is himself subject to random shocks which represent the noise the planner faces when computing the tradeoffs between individual utilities. Under the planner all agents internalize the effect of their choice on others. As a result, there's a unique average choice being selected. Secondly, it is shown that the choice selected without the planner is almost always socially inefficient, even when it has the same sign as the planner choice.

An application of statistical mechanics, from physicists rather than economists, is the field of econophysics. A review of much work is available in Castellano et al. (2009) which discusses statistical physics application to opinion, cultural and language dynamics as well as models of flocking and crowd behavior, hierarchy formation and spreading of social phenomena. The work of Bouchaud (2013) is of particular interests since it applies the stochastic Ising model to the study of market bubbles.

1.2.4 Quantal Response Equilibrium

A related notion is that of Quantal Response Equilibrium for boundedly rational agents. Players are assumed make mistakes as they play and the probability distribution of plays is then the QRE. This represents a radical shift from the notion of equilibria as points in the strategy space to equilibrium as probability distributions over all possible strategies. It is first introduced by McKelvey and Palfrey (1995) who proposes as an example the Logit Quantal Response Equilibrium, which is closely mirrored by the stationary distribution presented in this chapter. A recent application can be found in Kawagoe et al. (2018), which discusses QRE for Volunteer's dilemma and step public good provisions with binary decision.

Chapter 2

Preliminaries

2.1 Random Utility Models

Random Utility Models have a long history in the fields of economics and psychology, describing agents decision-making in the presence of uncertainty . They have numerous applications — product differentiation, labor economics, game theory, econometrics — and are the building block for describing agents behaviour in this thesis. This section provides a brief overview of random utility models. An in depth exploration of the history and techincal aspects of these models, as well as their application, is provided in Anderson et al. (1992).

For n alternative choices, an agent behaves according to a RUM when it's utility can be decomposed as

$$U_i = u_i + \epsilon_i, \ i = 1, \cdots, n.$$

Here u_i is the deterministic component of utility when choice i is picked. To each choice is associated an i.i.d. random variable, or a shock, ϵ_i with joint cumulative density $F(\epsilon_1, \cdots, \epsilon_n)$ defined over $\mathbb{R}^n$. The uncertainty, in the economics theory, is attributed to *modeller lack of informations*. That is to say, the point of view of the economist is that of an econometrician[1]. Manski (1977) lists four possible sources of uncertainty: non-observable characteristics, non-observable variations in individual utilities, measurement errors and functional missspecification. Psychology makes use of random utility models too, but interprets uncertainty as *fluctuations of personal preferences*. Regardless of the epistemologically different approaches, both have to

[1]Indeed, a lot of related work on interaction models focuses on conditions required to conduct inference. For an early perspective see McFadden (1981, 1984). More recent discussion can be found in Blume et al. (2015); Brock and Durlauf (2001).

the same technical implications.

The probability an agent picks the i-th choice can then be derived[2] following the principle of maximization of individual utilities: an agent will select i over j if, given the realization of the shocks, $U_i > U_j$, hence

$$P_i = \mathbb{P}\Big(U_i = \max_{j=1,\cdots,n} U_j\Big)$$

which rewrites as

$$P_i = \mathbb{P}(\epsilon_1 - \epsilon_i \leq u_i - u_1,\ \cdots,\ \epsilon_n - \epsilon_i \leq u_i - u_n).$$

For any possible realization x of ϵ_i the i-th alternative will be chosen with probability density $\prod_{i\neq j} F(u_i - u_j + x)$. Integrating over all possible values of x

$$P_i = \int_{\mathbb{R}} f(x) \prod_{i\neq j} F(u_i - u_j + x) dx. \tag{2.1}$$

where $f(x)$ is the density function of shocks. We are going to assume a specific form for errors, by specifiying that the ϵ_i follow a double exponential distribution (DED).

Definition 1 (Double Exponential Distribution) *A random variable X supported over $\mathbb{R}$ follows a double exponential distribution (or a Gumbel distribution or a type I Generalized Extreme Value distrbution) with scale parameter σ if its cumulative density function is given by*

$$F(x) = \mathbb{P}(\epsilon_i \leq x) = e^{-e^{-(\frac{x}{\sigma} - \gamma)}},$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant, and we write $X \sim DED(\sigma)$.

The main reason for using double exponential random variables is a practical one, as these distribution of error yields model where the selection probabilities have a (multinomial) logistic distribution [3] with a convenient closed form.

Definition 2 (Multinomial Logistic Distribution) *A discrete random variable Y supported over $\mathcal{I} = \{1, \ldots, n\}$ follows a multinomial logistic distribution with*

[2] More details on the derivation can be found in McFadden (1981, 1984). For a practical application see McFadden (1974) where RUM models are employed in the context of traffic prediction.

[3] Alternatively, models with normally distribute shocks could be employed. The jury is still out on which of the two type of error gives a better description of human behavior, see Kandori et al. (2008) for further discussion.

parameters $\zeta \in \mathbb{R}, \beta^{-1} > 0$ if the probability mass function of Y over $\mathcal{I}$ is

$$\mathbb{P}(Y = i) = \frac{e^{\beta u_i - \zeta}}{\sum_j e^{\beta u_j - \zeta}}$$

for some collection of positive weights $u_1, \ldots, u_n$ and we denote $Y \sim MLD(\zeta, \beta^{-1})$.

The connection between double exponential distribution of errors and multinomial logistic distribution is [4] given in the following theorem.

Theorem 1 (Holman and Marley) *Consider a random utility model where errors ϵ_i are $DED(\sigma)$, then choices for a given agent follow a multinomial logistic distribution with $\zeta = 0$ and $\beta^{-1} = \sigma$ and weights corresponding to the deterministic component of the utility.*

Proof:
The PDF of a double exponential is given by

$$f(x) = \frac{1}{\sigma} e^{-\left(\frac{x}{\sigma} - \gamma\right)} e^{-e^{-\left(\frac{x}{\sigma} - \gamma\right)}}.$$

From the definition of a double exponential,

$$F(u_i + x - u_j) = \exp\left[- \exp - \left(\frac{u_i + x - u_j}{\sigma} - \gamma \right) \right], \tag{2.2}$$

for $i, j \in \mathcal{I}$ and $i \neq j$. For readability use the change of variable $\varrho = \exp - \left(\frac{u_i + x - u_j}{\sigma} - \gamma \right)$ and $y_j = \exp \frac{u_j}{\sigma}$ and rewrite Eq. (2.1) as

$$\begin{aligned} P_i &= \int_0^\infty \exp(-\varrho) \prod_{i \neq j} \left[\exp\left(-\varrho \frac{y_j}{y_i} \right) \right] d\varrho \\ &= \int_0^\infty \exp\left[-\varrho \sum_{j=1}^n \frac{y_j}{y_i} \right] d\varrho \end{aligned} \tag{2.3}$$

The integration yields

$$P_i = -\frac{y_i}{\sum_{j=1}^n y_j} \exp\left[-\varrho \sum_{j=1}^n \frac{y_j}{y_i} \right]_0^\infty = \frac{y_i}{\sum_{j=1}^n y_j}.$$

which shows that the probability of choice of a single agent is logistic according to Definition 2 since $y_i = e^{\frac{u_j}{\sigma}}$ with $\zeta = 0$. □ As a corollary, we immediatly get the

[4] Attributed to Holman and Marley in Luce and Suppes (1965).

known fact that the difference between two $DED(\sigma)$ r.v. is logistic.

Corollary 1 (Difference of double exponential r.v.) *If $X, Y \sim DED(\sigma)$, then $X - Y$ is logistically distributed with variance $\frac{1}{\beta} = \sigma$ and zero mean.*

Proof: Consider the random utility model with $n = 2$. The probability of making choice 1 over choice 2 is

$$P_1 = \mathbb{P}(\epsilon_1 - \epsilon_2 \leq u_2 - u_1)$$

According to Theorem 1,

$$P_1 = \mathbb{P}(\epsilon_1 - \epsilon_2 \leq u_2 - u_1) = \frac{e^{\beta u_1}}{e^{\beta u_1} + e^{\beta u_2}} = \left[1 + e^{\beta(u_2 - u_1)}\right]^{-1}.$$

Let $\epsilon_1 = X$ and $\epsilon_2 = Y$ and $u_1 - u_2 = \Delta u$,

$$\mathbb{P}(X - Y \leq \Delta u) = \left[1 + e^{-\beta \Delta u}\right]^{-1},$$

the CDF of the logistic distribution, with β the inverse variance and zero mean. □

2.2 Markov Chains

This section recaps basic definitions on Markov chains and references some important tools used in the thesis. The notation employed is the standard one for Interacting Particle System followed in Liggett (1985).

2.2.1 Discrete Time Chains

A Markov chain in discrete time is a collection of random variable $\{\boldsymbol{x}_t\}_{t \geq 0}$, taking values $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \Lambda$ which is assumed finite. Elements of Λ are generically referred to as *states*. When they are vectors representing the state of a collection of particles or, in our case, the collection of choices of agents they are called *configurations*. In the following brackets denote probability measures. Bold letters denote states or configurations. Subscripts are used to denote intial conditions which are probability distribution over elements of the state space. When the subscript is a configuration it is assumed that the initial condition is given by the delta measure which assigns probability one to that configuration. This same notation is used for expectations. For example, $\mathbb{P}_\nu[\boldsymbol{x}_t = \boldsymbol{x}]$ is the probability of observing the chain $\boldsymbol{x}_t$ taking value $\boldsymbol{x} \in \Lambda$, given that the chain was started at $t = 0$ at some state whose probability is specified by the distribution ν. Similarly, $\mathbb{E}_{\boldsymbol{y}}[\boldsymbol{x}_t]$ is the expected value of the chain

at time t when the chain started in position $\boldsymbol{y}$ with probability one. The basic definitions about Markov chains presented here follows Norris (1998).

Let ν be a probability distribution over Λ and let P be a $|\Lambda| \times |\Lambda|$ stochastic matrix, that is, a matrix such that,

$$P_{\boldsymbol{xy}} \geq 0 \text{ and } \sum_{\boldsymbol{y} \in \Lambda} P_{\boldsymbol{xy}} = 1.$$

Definition 3 (Markov Chain) *The collection of* Λ *valued random variables* $\{\boldsymbol{x}_t\}_{t \geq 0}$ *is Markov-*(ν, P) *if its law* $\mathbb{P}_\nu$*, for all* $\boldsymbol{y}, \boldsymbol{z} \in \Lambda$ *and all events* $H_{t-1} = \cap_{\ell=0}^{t-1} \{\boldsymbol{x}_\ell = \boldsymbol{y}_\ell\}$ *such that* $\mathbb{P}_\nu[H_{t-1} \cap \{\boldsymbol{x}_t = \boldsymbol{z}\}] \neq 0$*, satisfies*

i. $\mathbb{P}_\nu[\boldsymbol{x}_0 = \boldsymbol{y}] = \nu[\boldsymbol{y}]$

ii. $\mathbb{P}_\nu[\boldsymbol{x}_{t+1} = \boldsymbol{z} | H_{t-1} \cap \{\boldsymbol{x}_t = \boldsymbol{y}\}] = P_{\boldsymbol{yz}}$

So, a collection of random variables is a Markov chain with initial distribution ν and transition matrix P when the probability of it's first realisation is governed by the initial distribution and the probability of the next realisation of the chain conditional on it's past only depends on the last realisation and is equal to P. The following alternative definition is called the *Markov property* or *memoryless property*, and brings home the crucial characteristic of Markov chains: the next realisation of the chain only depends on the current realisation.

Theorem 2 (Markov Property) *The collection of random variables* $\{\boldsymbol{x}_t\}_{0 \leq t \leq T}$ *is Markov-*(ν, P) *if and only if for all* $t \geq 0$ *and* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t \in \Lambda$

$$\mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_t = \boldsymbol{y}_t] = \nu[\boldsymbol{y}_0] P_{\boldsymbol{y}_0 \boldsymbol{y}_1} \ldots P_{\boldsymbol{y}_t \boldsymbol{y}_{t-1}}. \tag{2.4}$$

Proof: Suppose $\{\boldsymbol{x}_t\}_{t \geq 0}$ is Markov-(ν, P), then

$$\begin{aligned}
\mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_t = \boldsymbol{y}_t] &= \mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0] \mathbb{P}[\boldsymbol{x}_t = \boldsymbol{y}_t | \boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_{t-1} = \boldsymbol{y}_{t-1}] \\
&= \mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0] \mathbb{P}[\boldsymbol{x}_1 = \boldsymbol{y}_1 | \boldsymbol{x}_0 = \boldsymbol{y}_0] \ldots \mathbb{P}[\boldsymbol{x}_t = \boldsymbol{y}_t | \boldsymbol{x}_0 = \boldsymbol{y}_0 \ldots \boldsymbol{x}_{t-1} = \boldsymbol{y}_{t-1}] \\
&= \nu(\boldsymbol{y}_0) P_{\boldsymbol{y}_1 \boldsymbol{y}_0} P_{\boldsymbol{y}_2 \boldsymbol{y}_1} \ldots P_{\boldsymbol{y}_t \boldsymbol{y}_{t-1}}
\end{aligned}$$

If eq. (2.4) holds, then

$$\sum_{\boldsymbol{y}_T} \mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_T = \boldsymbol{y}_T] = \mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_{T-1} = \boldsymbol{y}_{T-1}],$$

and by induction this is true for any t. In particular, for $t = 0$

$$\mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0] = \tilde{V}_{\underline{h},m+1}[\boldsymbol{y}_0],$$

which is i from Definition 3. From this follows that

$$\mathbb{P}[\boldsymbol{x}_{t+1} = \boldsymbol{y}_{t+1} | \boldsymbol{x}_0 = \boldsymbol{y}_0,\ \ldots,\ \boldsymbol{x}_t = \boldsymbol{y}_t] = \frac{\mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_{t+1} = \boldsymbol{y}_{t+1}]}{\mathbb{P}[\boldsymbol{x}_0 = \boldsymbol{y}_0, \ldots, \boldsymbol{x}_t = \boldsymbol{y}_t]} = P_{\boldsymbol{y}_t \boldsymbol{y}_{t+1}}$$

which is ii. □

Definition 4 (Expectations) *Given a function f taking values in Λ, one step conditional expectations are defined as*

$$\mathbb{E}_{\boldsymbol{x}}[f(\boldsymbol{x}_1)] = \sum_{\boldsymbol{y} \in \Lambda} \mathbb{P}_{\boldsymbol{x}}[\boldsymbol{x}_1 = \boldsymbol{y}] f(\boldsymbol{y}) = \sum_{\boldsymbol{y} \in \Lambda} P_{\boldsymbol{x}\boldsymbol{y}} f(\boldsymbol{y}) = P f(\boldsymbol{x})$$

and t step expectations are then given by

$$\mathbb{E}_{\boldsymbol{x}}[f(\boldsymbol{x}_t)] = P^t f(\boldsymbol{x})$$

Once a chain is started at 0, the probability of any state being reached at some later time t is given by its distribution.

Definition 5 (Distributions) *The distribution of the chain at time t started from ν is defined recursively as*

$$\nu_t[\boldsymbol{x}] := \mathbb{P}_\nu[\boldsymbol{x}_t = \boldsymbol{x}] = \nu_{t-1} P_{\boldsymbol{x}} = \sum_{\boldsymbol{y} \in \Lambda} \nu_{t-1}[\boldsymbol{y}] P_{\boldsymbol{y}\boldsymbol{x}},$$

or equivalently in vector notation

$$\nu_t = \nu_{t-1} P,$$

meaning that

$$\nu_t = \nu_0 P^t.$$

How does the chain behave after a long time? Or, in other words, what does the distribution ν_t looks like in the limit $t \to \infty$? When this limit exist it is referred to as the stationary distribution, describing the long run behavior of the chain.

Definition 6 (Stationary Distribution) *A probability distribution π on Λ is called stationary if for all $\boldsymbol{x} \in \Lambda$*

$$\pi[\boldsymbol{x}] = \sum_{\boldsymbol{y} \in \Lambda} \pi[\boldsymbol{y}] P_{\boldsymbol{y}\boldsymbol{x}},$$

or equivalently in vector notation

$$\pi = \pi P.$$

If π is stationary and $\nu_0 = \pi$ then $\nu_t = \pi$.

A particular class of chain are those chains for which any state can always be reached starting from any other state.

Definition 7 (Irreducibile chain) *A Markov chain is said to be irreducibile if for any pair of states $\boldsymbol{x}, \boldsymbol{y} \in \Lambda$ there exists a $t \geq 0$ such that*

$$\mathbb{P}[\boldsymbol{x}_{t+1} = \boldsymbol{y} | \boldsymbol{x}_t = \boldsymbol{x}] = P^t_{\boldsymbol{x}\boldsymbol{y}} > 0.$$

Irreducibile chains over finite state space always possess a unique stationary distribution. In order to prove this we need to first define harmonic functions. Locally harmonic functions are such that their value at a point is equal to the average around that point and play an important role in the study of Markov chains.

Definition 8 (Harmonic functions) *A function $h : \Lambda \to \mathbb{R}$ is harmonic at $\boldsymbol{x} \in \Lambda$ if*

$$h(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Lambda} P_{\boldsymbol{x}\boldsymbol{y}} h(\boldsymbol{y}).$$

If h is harmonic at all elements of $\Lambda = \{\ldots \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots\}$ then it is harmonic on Λ, meaning that $Ph = h$, where $h^T = (\ldots, h(\boldsymbol{x}), h(\boldsymbol{y}), h(\boldsymbol{z}), \ldots)$ is a column vector.

Lemma 1 *If a chain is irreducibile and h is harmonic on Λ, then h is a constant function.*

Proof: Since Λ is finite, there exists an $\boldsymbol{x} \in \Lambda$ such that $\boldsymbol{x} = \arg\max_{\boldsymbol{y} \in \Lambda} h(\boldsymbol{y})$.

Pick a $\boldsymbol{z} \in \Lambda$ such that $P_{\boldsymbol{xz}} > 0$ and assume $h(\boldsymbol{x}) > h(\boldsymbol{z})$. By harmonicity at $\boldsymbol{x}$,

$$\begin{aligned} h(\boldsymbol{x}) &= \sum_{\boldsymbol{y}\in\Lambda} P_{\boldsymbol{xy}} h(\boldsymbol{y}) \\ &= P_{\boldsymbol{xz}} h(\boldsymbol{z}) + \sum_{\boldsymbol{y}\neq\boldsymbol{z}} P_{\boldsymbol{xy}} h(\boldsymbol{y}) \\ &\leq P_{\boldsymbol{xz}} h(\boldsymbol{z}) + \sum_{\boldsymbol{y}\neq\boldsymbol{z}} P_{\boldsymbol{xy}} h(\boldsymbol{x}) \\ &< P_{\boldsymbol{xz}} h(\boldsymbol{x}) + \sum_{\boldsymbol{y}\neq\boldsymbol{z}} P_{\boldsymbol{xy}} h(\boldsymbol{x}) \\ &= \left(\sum_{\boldsymbol{y}\in\Lambda} P_{\boldsymbol{xy}} \right) h(\boldsymbol{x}) \\ &= h(\boldsymbol{x}) \end{aligned}$$

The strict inequality rests on the fact that $P_{\boldsymbol{xz}} > 0$ and the assumption that $h(\boldsymbol{x}) > h(\boldsymbol{z})$ and leads to the contradiction that $h(\boldsymbol{x}) > h(\boldsymbol{x})$. Hence, it must be the case that $h(\boldsymbol{x}) = h(\boldsymbol{z})$. Since the chain is irreducibile, then there is a path from $\boldsymbol{x}$ to $\boldsymbol{y}$, let it be $\boldsymbol{x}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \boldsymbol{y}$ where each step has non-zero probability. So, for example $P_{\boldsymbol{xv}} > 0$ and therefore $h(\boldsymbol{x}) = h(\boldsymbol{v})$. Iterating this logic on the whole path makes as conclude that

$$h(\boldsymbol{x}) = h(\boldsymbol{v}) = \cdots = h(\boldsymbol{y}).$$

So $h(\boldsymbol{x}) = h(\boldsymbol{y})$ for all $\boldsymbol{y} \in \Lambda$ and h is therefore constant. □

Theorem 3 (Uniqueness of stationary distribution) *If a chain is irreducibile and has a stationary distribution π, the stationary distribution is unique.*

Proof: By Lemma 1, the only solution to $Ph = h$ are of the form $h^T = c(1 \cdots 1)^T$. Thus, $(P - I)h = 0$ implies that the dimension of the null space of $P - I$ is 1. By the rank-nullity theorem, $(P - I) = |\Lambda| - 1$. Taking the transpose of this is $(P - I) = (P - I)^T = (P^T - I) = |\Lambda| - 1$ and again by the rank-nullity theorem, the null space of $P^T - I$ is one. This means that the set of vectors $v \in \mathbb{R}^{|\Lambda|}$ that solve $(P^T - I)v = 0$ has also dimension one. All solutions have the form $Pv = \lambda\pi$ for some scalar . But for v to be a distribution, $\lambda = 1$, so that $v = \pi$. □

Another crucial property of a Markov chain is reversibility with respect to some distribution, which is particularly helpful in identifying stationary distributions.

Definition 9 (Reversibility) *A Markov chain is said to be reversibile with respect*

to some measure μ if

$$\mu[\boldsymbol{x}]P_{\boldsymbol{xy}} = \mu[\boldsymbol{y}]P_{\boldsymbol{yx}}.$$

Indeed, if a chain is irreducibile the reversible measure is the unique stationary distribution of the chain since summing over $\boldsymbol{y}$ on both sides

$$\mu[\boldsymbol{x}] = \sum_{\boldsymbol{y}} \mu[\boldsymbol{y}]P_{\boldsymbol{yx}} \tag{2.5}$$

returns the definition of stationarity.

2.2.2 Expected Hitting Times

Given a Markov chain $\{\boldsymbol{x}_t\}_{t\geq 0}$ with transition matrix P, the first hitting time of a subset A of Ω is the random variable $\mathcal{T}_A : \Omega \to \mathbb{N}$ given by

$$\mathcal{T}_A = \inf\{t \geq 0 : \boldsymbol{x}_t \in A\}$$

with the convention that the infimum of the empty set is given by ∞. The mean hitting time for $\{\boldsymbol{x}_t\}_{t\geq 0}$ to reach A when the process starts in $\boldsymbol{x}$ (or the initial distribution is $\delta_{\boldsymbol{x}}$) is given by

$$\tau_{\boldsymbol{x}A} = \mathbb{E}_{\boldsymbol{x}}[\mathcal{T}_A] = \sum_{t\in\mathbb{N}} t\mathbb{P}_{\boldsymbol{x}}[\mathcal{T}_A = t] + \infty\mathbb{P}_{\boldsymbol{x}}[\mathcal{T}_A = \infty]$$

The following theorem, due to Bovier and den Hollander (2015), relates the expected hitting time of Markov chain over the line to its stationary distribution

Theorem 4 (Bovier) *Consider a Markov chain m_t over the line segment Γ with nearest neighbor transitions and with stationary measure μ. The expected hitting time for the chain started at a to reach b is given by*

$$\mathbb{E}_a[\tau_b] = \sum_{\substack{m,m'\in\Gamma \\ m\leq m' \\ b\leq m\leq a}} \frac{\mu[m]}{\mu[m']} \frac{1}{P_{mm-1}} \tag{2.6}$$

2.2.3 Lumping

The state space of a Markov chain can be reduced[5] by partitioning its state space and accordingly its transition probabilities. This is very useful both to gain intuition and for computational speed up.

[5] See Kemeny and Snell (1960) for a detailed discussion of lumping.

Definition 10 (Lumped) *Consider a discrete-time Markov chain* $\{\boldsymbol{x}_t\}_{t\geq 0}$ *with transition matrix* P *and state space* Λ*. Given a partition* $\Gamma = \{\Gamma_1, \ldots, \Gamma_k\}$ *of the state space we define the lumped process* $\{m_t\}_{t\geq 0}$ *with state space* Γ *and transition matrix* $R_{\Gamma_j\Gamma_\ell} = \sum_{\boldsymbol{x}\in\Gamma_j}\sum_{\boldsymbol{y}\in\Gamma_\ell} P_{\boldsymbol{xy}}$*.*

Unfortunately, the process obtained in this way need not be itself Markov. The following theorem gives necessary and sufficient condition for lumpability.

Theorem 5 (Lumpable Markov Chain) *Consider a lumped process* $\{m_t\}$ *with transition matrix* R *over a partition* Γ *of the original state space. The lumped chain is Markov if and only if for any* Γ_j, Γ_i *it holds that*

$$\sum_{\boldsymbol{z}\in\Gamma_j} P_{\boldsymbol{xz}} = \sum_{\boldsymbol{z}\in\Gamma_j} P_{\boldsymbol{yz}}, \text{ for all } \boldsymbol{x}, \boldsymbol{y} \in \Gamma_i. \tag{2.7}$$

What this means, is that the probability of reaching an element Γ_j of the partitioned space should be constant whenever starting from elements of the original space that belong to the same partition element Γ_i.

2.2.4 Metastability

Definition 11 (Metastability) *A family of Markov chain indexed by* N *is called metastable if there exists a collection of disjoint sets* $B_i \subset \Omega$*, such that*

$$\frac{\sup_{m\notin\cup_i B_i}\mathbb{E}_m[\tau_{\cup_i B_i}]}{\inf_i \inf_{m\in B_i}\mathbb{E}_m[\tau_{\cup_j B_j\setminus B_i}]} = o(1), \ N\to\infty. \tag{2.8}$$

Meaning that the longest time to enter any metastable set is much smaller than the time to leave the least stable one and increasingly so as $N \to \infty$.

2.2.5 Laplace-Varadhan Lemma

The Laplace-Varadhan lemma is a result that comes into play in many forms in large deviation theory. The gist is that it allows computation of limit integrals of the form $\int e^{Nf(x)}dx$, by ignoring suitably small terms and focusing instead on the point where most of the mass of the integral is centered. This type of calculations are also referred as saddle point techniques. When x is a random variable the integral is an expectation of a particular function, making it particularly useful in the field of probability. Here we give a statement of the lemma which will later be used to approximate expected hitting times of a Markov chain. The following has been given to me in the current form by Dr. Massimo Iberti, for more detail see Lemma 6.2 in Bovier and den Hollander (2015).

Lemma 2 (Laplace-Varadhan Lemma) *As N goes to infinity*

$$\frac{1}{N}\ln\sum_{i=0}^{N} e^{Nf(i/N)}\mathbb{1}_{a\le i/N\le b} \to \max_{x\in[a,b]} f(x).$$

hence, for a finite N,

$$\frac{1}{N}\ln\sum_{i=0}^{N} e^{Nf(i/N)}\mathbb{1}_{a\le i/N\le b} = \max_{x\in[a,b]} f(x) + \mathcal{O}(\ln N/N).$$

Proof:

$$\frac{1}{N}\ln\sum_{i=0}^{N} e^{Nf(i/N)}\mathbb{1}_{a\le i/N\le b} \ge \max_{i:a\le i/N\le b} f(i/N) \to \max_{x\in[a,b]} f(x).$$

$$\begin{aligned}\frac{1}{N}\ln\sum_{i=0}^{N} e^{Nf(i/N)}\mathbb{1}_{a\le i/N\le b} &\le \frac{1}{N}\ln\left[\Big|\{i : a\le i/N\le b\}\Big| \max_{i:a\le i/N\le b} e^{f(i/N)}\right]\\ &\le \frac{\ln N|b-a|}{N} + \max_{i:a\le i/N\le b} f(i/N) \to \max_{x\in[a,b]} f(x).\end{aligned}$$

2.2.6 Coupling

Coupling is a technique employed in probabilistic proofs. It consists of constructing a joint distribution between two stochastic processes whose marginal distributions are that of the original processes. The following is a definition for any random variable which extends trivially to stochastic processes. A coupling of two probability distribution μ and ν, consists of random variables (X, Y) over a single probability space with a joint probability measure π such that the marginal distribution of X is μ and the marginal distribution of Y is ν.

Definition 12 (Coupling) *Given two probability distributions μ, ν over the same probability space $\mathcal{P}(\Omega)$, a coupling consists of a joint probability distribution $\pi \in \mathcal{P}(\Omega\times\Omega)$, such that*

$$\begin{aligned}&\pi[\{(x,y)\in\Omega\times\Omega : x\in A\}] = \mu[A] \text{ and}\\ &\pi[\{(x,y)\in\Omega\times\Omega : y\in B\}] = \nu[B],\end{aligned} \tag{2.9}$$

where A and B are measurable subsets of Ω.

Notice that the choice of the joint distribution is not uniquely specified by

the marginals. The trick to construct a useful coupling is therefore to pick a joint distribution with nice properties. For us, the property we are interested in will be stochastic monotonicity.

Definition 13 (Stochastic Monotonicity) *A partial order over $\mathcal{P}(\Omega)$ is given by*

$$\mu \leq \nu \text{ if } \mathbb{E}^{\mu}[f] \leq \mathbb{E}^{\nu}[f] \tag{2.10}$$

for all increasing and continuous functions f over Ω.

The following theorem is the workhorse of coupling.

Theorem 6 (Strassen) *Suppose $\mu, \nu \in \mathcal{P}(\Omega)$, then $\mu \leq \nu$ if and only if there exists a successful coupling $\pi \in \mathcal{P}(\Omega \times \Omega)$, that is, a coupling such that, for all Borel sets in Ω,*

a) $\pi\{(x, y) : x \in A\} = \mu[A]$

b) $\pi\{(x, y) : y \in A\} = \nu[A]$

c) $\pi[\{(x, y) : x \leq y\}] = 1$

Proof: Omitted because of no direct relevance for this thesis. The interest reader can find it as proof of Theorem 2.4 in Liggett (1985), page 24. □

Strassen's theorem is useful because it allows us to verify that two measures, and therefore two expectations by Def. 13, are stochastically monotone.

2.3 Stochastic Optimization

This section introduces the notation and the fundamental concepts of Stochastic Dynamic Programming, the main reference is *Introduction to Dynamic Stochastic Programming*, Ross (1983).

2.3.1 Finite Stage Problems

Imagine that some decision maker observes a Markov chain for n steps. At each time step, after the chain has realized, the decision maker can select an action h, after which a cost $C(i, h)$ accrues depending only on the current state of the chain and action chosen. A new state is then realized with probability $P_{ij}(h)$. Goal of the decision maker is then to select the action which minimizes the expected total costs

accrued over the n stages. Define $V_n(i)$ to be the minimum expected cost for a n stage problem when the Markov chain starts in state i, clearly

$$V_1(i) = \min_h C(i,h)$$

so that the one stage optimal policy is to pick $\bar{h} = \arg\min C(i,h)$ when state i is observed. Whene there are n stages, after the $n-th$ stage has realized, the next state j is realized with probability $P_{ij}(h)$ and so we obtain the following recursive definition known as the optimality equation:

$$V_n(i) = \min_h \left[C(i,h) + \sum_j P_{ij}(h) V_{n-1}(j) \right].$$

This functional equation can be solved recursively for the n stage optimal policy, albeit it becomes computationally expensive quickly. But it can also be used to obtain structural information on the behavior of the optimal policy, and most importantly, it is the foundation upon which discounted dynamic programming can be built.

2.3.2 Discounted Dynamic Programming

Denote X_t a Markov chain taking value in some countable state space Γ indexed by non negative integers. After each realization of the chain an action must be chosen, let H denote the finite set of all possible actions. When the process is in state i and action h is chosen, then:

- A cost $C(i,h)$ is paid,
- The Markov chain next state is realized with probability $P_{ij}(h)$.

Note that the second assumption is equivalent to assuming that

$$\mathbb{P}[X_{t+1} = j \mid X_0,\ h_0,\ \ldots,\ X_t = i,\ h_t = h] = P_{ij}(h) \tag{2.11}$$

so that both the cost and the transition probabilities are function only of the previous state and the action chosen. It is also assumed that the costs are bounded by some constant K. A rule to choose actions is called a policy on which no restriction is imposed a priori. Any rule that allows to chose an action, even when dependent on the history of the process or when it selects an action randomly with a certain probability, is a valid policy.

Definition 14 (Stationary policy) *A policy is called stationary if it is non-random and the action it chooses only depends on the state of the process at times t.*

A stationary policy is therefore a function $f : \Gamma \to H$ mapping states into actions. When a stationary policy is applied to choose the action then the chain X_t is again Markov with transition probabilities $P_{ij}(f(i))$, motivating the name of Markov decision process. We shall restrict our attention to the class of stationary policies. Define the expected total discounted cost under policy π as

$$V_\pi(i) := \mathbb{E}\left[\sum_{t\geq 0} \delta^t C(X_t, h_t)|X_0 = i\right] \tag{2.12}$$

which is well defined because costs are bounded and $0 < \delta < 1$.

2.3.3 Optimality equation

The value function is defined as the function that associates to each state the minimal expected total discounted cost :

$$V_i = \inf_\pi V_\pi(i). \tag{2.13}$$

The following theorem yields a functional equation which the value function must satisfy.

Theorem 7 (Optimality Equation)

$$V_i = \min_h \left[C(i,h) + \delta \sum_j P_{ij}(h)V_j\right]. \tag{2.14}$$

Proof: Consider an arbitrary policy π, and assume that it chooses action h with some probability p_h. Then,

$$V_\pi(i) = \sum_h p_h \left[C(i,h) + \sum_j P_{ij}(h)V_\pi^1(j)\right],$$

with $V_\pi^1(j)$ representing expected discounted costs from $t = 1$ onward, when policy π is used and j is the current state of the process. But then this is the same as if

the process had started in j, but all costs are discounted by δ, meaning that

$$\begin{aligned} V_\pi^1(j) &= \mathbb{E}\left[\sum_{t\geq 1}\delta^t C(X_t,h_t)|X_0=i\right] \\ &= \delta\mathbb{E}\left[\sum_{t\geq 0}\delta^t C(X_t,h_t)|X_0=i\right] \\ &= \delta V_\pi(j) \geq \delta V_j,, \end{aligned}$$

hence

$$\begin{aligned} V_\pi(i) &\geq \sum_h p_h\left[C(i,h)+\delta\sum_j P_{ij}(h)V_j\right] \\ &\geq \sum_h p_h \min_h\left[C(i,h)+\delta\sum_j P_{ij}(h)V_j\right] \\ &= \min_h\left[C(i,h)+\delta\sum_j P_{ij}(h)V_j\right]. \end{aligned}$$

But π is arbitrary, therefore,

$$V_i \geq \min_h\left[C(i,h)+\delta\sum_j P_{ij}(h)V_j\right]. \tag{2.15}$$

Having given a lower bound, we can now show that the upper bound is the same by picking h_0 so that

$$C(i,h_0)+\delta\sum_j P_{ij}(h)V_j = \min_h\left[C(i,h)+\delta\sum_j P_{ij}(h)V_j\right]$$

and let π be a policy that picks h_0 in $t=0$ and after that picks a policy $\tilde{\pi}$ such that $\mathbb{V}_{\tilde{\pi}} \leq V_j+\epsilon$. It follows that

$$\begin{aligned} V_\pi(i) &= C(i,h_0)+\delta\sum_j P_{ij}(h_0)V_{\tilde{\pi}}(j) \\ &\leq C(i,h_0)+\delta\sum_j P_{ij}(h_0)V_j+\delta\epsilon, \end{aligned}$$

but given that $V_j \leq V_\pi$, we get

$$V_j \leq C(i, h_0) + \delta \sum_j P_{ij}(h_0) V_j + \delta\epsilon. \tag{2.16}$$

The results follow from (2.15) and (2.16) since ϵ is arbitrary $\square$.
The next theorem shows that the policy satisfying the optimality equation is optimal.

Theorem 8 *Let f be a stationary policy such that*

$$C(i, f(i)) + \delta \sum_j P_{ij}(f(i)) V_j = \min_h \left[C(i, h) + \delta \sum_j P_{ij}(h) V_j \right] \tag{2.17}$$

then

$$V_f(i) = V_i \text{ for all } i, \tag{2.18}$$

therefore f is optimal.

Proof: Using (2.14) we have

$$V_i = \min \left[C(i, h) + \delta \sum_j P_{ij}(h) V_j \right] = C(i, f(i)) + \delta \sum_j P_{ij}(f(i)) V_j \tag{2.19}$$

which is equivalent to the two-stage version of the problem where we use policy f in the first stage and obtain a terminal reward V_j for some j. But then, the terminal reward is the same as using the policy for one more stage and gaining another terminal reward. Repeating this argument we can write

$$V_i = \mathbb{E}[\text{n-stage cost using } f | X_0 = i] + \delta^t \mathbb{E}[V_{X_t} | X_0 = i].$$

As $t \to \infty$ using $0 < \delta < 1$ and the fact that $V_i < \frac{K}{1-\delta}$ we get the statement $\square$.
Clearly we would like the solution of the optimality equation to be unique. This is the statement of the following theorem.

Theorem 9 *V is the unique bounded solution of the optimality equation* (2.14).

Proof: suppose that $u(i)$ is a bounded function that satisfies the optimality equation

$$u(i) = \min \left[C(i, h) + \delta \sum_j P_{ij}(h) u(j) \right]$$

Let $\bar{h}$ be such that $u(i) = \left[C(i,\bar{h}) + \delta \sum_j P_{ij}(\bar{h})u(j)\right]$ then since V_i satisfies the optimality equation

$$\begin{aligned} u(i) - V(i) &= C(i,\bar{h}) + \delta \sum_j P_{ij}(\bar{h})u(j) - \min_h [C(i,h) + \delta \sum_j P_{ij}(h)V_j] \\ &\geq \delta \sum_j P_{ij}(\bar{h})[u(j) - V_j] \\ &\geq \delta \sum_j P_{ij}(\bar{h})|u(j) - V_j| \\ &\geq \delta \sum_j P_{ij}(\bar{h}) \inf_j |u(j) - V_j| \\ &= \delta \inf_j |u(j) - V_j|. \end{aligned}$$

Inverting $u(i)$ and $V(i)$

$$V(i) - u(i) \leq \delta \inf_j |V_j - u(j)|$$

which means that

$$|V(i) - u(i)| \leq \delta \inf_j |V_j - u(j)|$$

and since this holds for all i it holds for the infimum

$$\inf_i |V(i) - u(i)| \leq \delta \inf_j |V_j - u(j)|$$

and therefore, given that $\delta < 1$,

$$\inf_i |V(i) - u(i)| = 0. \ \square$$

2.3.4 Successive approximation

From theorem 2.14 we know that if V where known then we could find the optimal policy as the action, for each state, that minimizes

$$C(i,h) + \delta \sum_j P_{ij}(h)V_j$$

The function V can be obtained as a limit of the n stage problem, as illustrated by the following theorem.

Theorem 10 *Define the value of a one stage problem started in state i as*

$$V_1(i) = \min_h \left[C(i,h) + \delta \sum_j P_{ij}(h) V_0(j) \right]. \tag{2.20}$$

where V_0 is an arbitrary bounded function. Let, for $n > 1$

$$V_n(i) = \min_h \left[C(i,h) + \delta \sum_j P_{ij}(h) V_{n-1}(j) \right], \tag{2.21}$$

denote the value of a n stage problem, then

1. *If $V_0 \equiv 0$ then $|V_0 - V_n| \leq \delta^{n+1} \frac{K}{1-\delta}$.*
2. *For any bounded V_0, $V_n \to V$ uniformly.*

Proof: To being note that given any policy

$$\begin{aligned} &|\mathbb{E}\left[\text{Costs from time } (n+1) \text{ onward} | X_0 = i\right]| \\ &= \left| \mathbb{E}\left[\sum_{t=n+1}^{\infty} \delta^t C(X_t, h_t) | X_0 = i \right] \right| \\ &\leq \frac{\delta^{n+1} K}{1-\delta} \end{aligned}$$

Suppose that $V_0 \equiv 0$, then V_n is the minimal expected cost for an n stage problem with terminal cost V_0. Then, for some optimal policy $\bar{h}$

$$\begin{aligned} V(i) &= \mathbb{E}_{\bar{h}}[\text{Costs in the first n-stages}] + \mathbb{E}_{\bar{h}}[\text{Costs in subsequent stages}] \\ &\geq V_n(i) - \frac{\delta^{n+1} K}{1-\delta}. \end{aligned}$$

To go the other direction note that V must be smaller than the expected costs of the policy that uses the n-stage optimal policy for n stages and any arbitrary policy for the rest of the time. Hence,

$$\begin{aligned} V(i) &\leq V_n(i) + \mathbb{E}[\text{Costs in subsequent stages}] \\ &\leq V_n(i) + \frac{\delta^{n+1} K}{1-\delta}, \end{aligned}$$

which proves the first statement. To prove the second, we let V_n^0 denote V_n when

$V_0 \equiv 0$. Let h_0 denote the action such that

$$V_1(i) = \min_h \mathbb{E}\left[C(i,h) + \delta \sum_j P_{ij}(h) V_0(j)\right] = \mathbb{E}\left[C(i,h_0) + \delta \sum_j P_{ij}(h_0) V_0(j)\right].$$

From the defintion of V_n^0

$$V_1^0(i) = \min_h \mathbb{E}\left[C(i,h) + 0\right]$$

which imply

$$|V_1(i) - V_1^0| \geq \left|\mathbb{E}\left[C(i,h_0) + \delta \sum_j P_{ij}(h_0) V_0(j)\right] - \mathbb{E}[C(i,h_0)]\right| \geq \delta \inf_j |V_0(j)|.$$

Iterating this reasoning we obtain

$$|V_n(i) - V_n^0(i)| \geq \inf_j \delta^n |V_0(j)|.$$

□

2.3.5 Policy Improvement Algorithm

Once V is determined the optimal policy chooses h to minimize $C(i,h)+\delta \sum_j P_{ij}(h)V_j$. Consider some stationary policy g for which we have computed the expected costs V_g. Now consider a policy f that minimizes $C(i,h) + \delta \sum_j P_{ij}(h)V_f(j)$. How good is f compared to g? It turns out that f is at least as good as g and if it is not strictly better for at least one initial state, then g and f are both optimal, yielding a strategy to obtain g and V computationally.

Theorem 11 *Let g be a stationary policy with expected costs V_g and let f be a policy such that*

$$C(i,f(i)) + \delta \sum_j P_{ij}(f(i)) V_g(j) = \min_h \left[C(i,h) + \delta \sum_j P_{ij}(h) V_g(j)\right], \qquad (2.22)$$

then

$$V_f(i) \leq V_g(i), \text{ for all } i,$$

and if $V_f(i) = V_g(i)$ for all i, then $V_g = V_f = V$.

Proof: Since

$$\min_h \left[C(i,h) + \delta \sum_j P_{ij}(h) V_g(j) \right] \leq C(i,g(i)) + \delta \sum_j P_{ij}(g(i)) V_g(j) = V_g(i)$$

then equation (2.22) implies

$$C(i,f(i)) + \delta \sum_j P_{ij}(f(i)) V_g(j) \leq V_g(i), \text{ for all } i. \tag{2.23}$$

This inequality states that using f for one stage and g afterward is better than using g in all stages. But this argument is valid for all stages, proving that $V_f(i) \leq V_g(i)$. If we suppose that $V_f(i) = V_g(i)$ for all i

$$C(i,f(i)) + \delta \sum_j P_{ij}(f(i)) V_g(j) = V_f(i),$$

then substituting V_g with V_f in (2.22) we get

$$V_f(j) = \min_h \left[C(i,h) + \delta \sum_j P_{ij}(h) V_f(j) \right], \tag{2.24}$$

meaning that V_f satisfies the optimality equation and by uniqueness (Thm. 9), we conclude $V_f = V$. □

In practice, the policy improvement algorithm gives us a computational tool to obtain the optimal policy. We summarize it using vector notation. First, fix an $\epsilon > 0$ which will serve as a stopping criteria. Pick any arbitrary policy π, and compute

$$V_\pi = C_\pi + \delta P_\pi V_\pi,$$

that is, solve the linear system $V_\pi = (I - \delta P_\pi)^{-1} C_\pi$. Using the newly computed values of V_π propose a new policy that satisfies

$$\omega = \arg\min_h C_h + \delta P_h V_\pi.$$

Set $\pi = \omega$ and repeat until $|V_\omega - V_\pi| \leq \epsilon$, at which point ω is the optimal policy.

Chapter 3

Setup

In our model a large population of agents holds one of two possible positions or beliefs. The vector of agents positions at time t is referred to as a *configuration*. At every time step one random agent revises its choice following their utility function which depends on the current configuration, the policy choice of an external planner and some idiosyncratic preference shock.

Agents continuously receive shocks and since their choice only depend on the current configuration the model can be cast as Markov chain over the set of all possible configurations which constitues the *state space* of the chain. The "solution" is then given by the probability of observing a configuration in the long term given by the stationary distribution of the chain. The chain will be shown to exhibit *long lived equilibria*, region of the state space where the chain spends a disproportionately large amount of time, known as *metastable sets* in the Markov chain literature.

Previous work sharing similar setup focused on which of the many possible configurations would be most likely: in (Durlauf, 1996; Blume, 1993; Kandori et al., 2008) agents play a one shot coordination game and the model is solved in the limit of an infinite population, possibily with a multiplicity of equilibria. In a different setup, but similar spirit, (Young, 1993) agents' play a coordination game with limited memory and making random mistakes eventually converging on a multiplicity of equilbria. The problem of equilibrium selection is tackled by showing that in the limit of small mistakes the stochastically dominating equilibrium is selected. Instead of asking which equilibrium survives, this thesis is concerned with the questions: *how do agents transition from one equilibrium to the other? How long does it take? How is this affected by the policy chosen by an external planner?*.

This chapter is devoted to setting up the model in terms of a Markov chain and study its long term behavior, borrowing heavily from the field of Statistical

Mechanics and Interacting Particle Systems. The most important reference is the summary of the Curie-Weiss model in Chapter 13 of Bovier and den Hollander (2015), which is used as a template for all calculations in this chapter, albeit details between the model differ requiring new proofs for several statements.

In this chapter it is assumed that the planner policy is fixed. The following chapters are dedicated to the planner problem asking what the optimal policy should be. Proofs of all propositions are in section 3.6.

A note on time

From now on we work in the framework of discrete time Markov chain. This greatly simplifies all computation, but it discards information on how long agents take to activate and revise their choice. This can be easily recovered later on since, due to the Markov property, it can be shown that waiting time and specific realisations of a Markov chain are independent. In other word, it is always possible to simulate a continuous time Markov chain by first drawing a realisation from its discrete analogue and after that simulate the expected *waiting* times spent in each state[1].

3.1 Timing and utility

There are N agents and each one can take up one of two positions $x_i \in \{-1, 1\}$. A configuration, a vector of all agents choices, is denoted $\boldsymbol{x} \in \{-1, 1\}^N$ and $\boldsymbol{x}_t$ when necessary to specify a timestep $t \in \mathbb{N}_+$. When agents are called to revise their choice they do so according to the utility function:

$$U_i(\boldsymbol{x}) = hx_i - \frac{\gamma}{4N} \sum_{j \neq i} (x_i - x_j)^2 + \epsilon_{x_i}, \tag{3.1}$$

where $h \in [-1, 1]$ is the policy chosen by some external planner and taken as given by agents. The summation expresses the fact that agents like to coordinate with other agents, how much so depends[2] on the strength of the coordination motive $\gamma \in [0, 1]$. This means that the agents, ignoring for the moment the idiosyncratic preference shock, increase their utility by choosing the sign of x_i that either aligns with the average choice of other agents or with the planner policy, whichever is higher. Preference shocks ϵ_{x_i}, one for each possible choice, follow a generalised

[1] See Kobayashi et al. (2011), Theorem 16.2, page 461.

[2] An alternative interpretation is that this parameter represents how often, on average, agents interact with each other or how often agents investigate the average population choice before making a decision.

double exponential distribution with zero mean and scale parameter $1/\beta$ given in Definition 1. Hence, agents follow a random utility model with two possible choices.

At each time step t one agent is selected uniformly at random, preference shocks are drawn and the agent myopically maximizes their utility by choosing $x_i \in \{-1, 1\}$, taking as given the policy choice h in period t from the planner and the choices of all other agents.

For a given configuration $\boldsymbol{x}$ we denote $\boldsymbol{x}^i$ the configuration where the $i-th$ agent has changed their choice $x_i = -x_i^i$, then the probability of switching choice is given, according to the corollary of Theorem 1, by

$$\begin{aligned} P_{\boldsymbol{x}\boldsymbol{x}^i} &= \mathbb{P}\left[U_i(\boldsymbol{x}^i) > U_i(\boldsymbol{x})\right] \\ &= \mathbb{P}\left[\epsilon_{x_i} - \epsilon_{x_i^i} < h(x_i^i - x_i) + \frac{\gamma}{N}\left(x_i^i \sum_{j\neq i} x_j^i - x_i \sum_{j\neq i} x_j\right)\right] \\ &= \left\{1 + e^{-\beta[\overline{U}_i(\boldsymbol{x}^i) - \overline{U}_i(\boldsymbol{x})]}\right\}^{-1} \end{aligned} \tag{3.2}$$

where $\bar{U}_i = hx_i - \frac{\gamma}{2N}\sum_{j\neq i}(x_i - x_j)^2$ is the non-stochastic component of the utility function. Equation (3.2) states that the probability of an agent changing their current choice is proportional to the difference in utilities multiplied by β , the inverse of the variance of the preference shocks. This allows us to interpret β as rationality parameter. Large β means shocks are tiny and agents will be most likely to revise their choice if it is rational to do so, i.e. if the difference in utility is positive. Conversely, for a small β, shocks[3] will be large and agents tend to commit a mistake more often. In the limit of $\beta \to \infty$ we recover rational agents, while for $\beta \to 0$ the choice becomes completely random. We will assume in the following that $\beta > 1$, guaranteeing a modicum of rationality.

Given the update dynamics described above $\boldsymbol{x}_t$ is a Markov chain over the set of possible configurations $\Lambda_N = \{-1, 1\}^N$ with transition matrix P. For any two configuration $\boldsymbol{x}, \boldsymbol{y} \in \Lambda_N$ the elements of P are

$$P_{\boldsymbol{x}\boldsymbol{y}} = \begin{cases} 0, & ||\boldsymbol{x} - \boldsymbol{y}|| > 2 \\ \frac{1}{N} P_{\boldsymbol{x}\boldsymbol{x}^i}, & ||\boldsymbol{x} - \boldsymbol{y}|| = 2 \\ 1 - \frac{1}{N}\sum_i P_{\boldsymbol{x}\boldsymbol{x}^i}, & ||\boldsymbol{x} - \boldsymbol{y}|| = 0 \end{cases}. \tag{3.3}$$

[3]Interpretation of these types of shock is discussed extensively in (Anderson et al., 1992; Blume et al., 2015).

3.2 Stationary Distribution

The formulation of our model ensures that agents, starting from any single configuration, will revise their choices moving toward their preferred choice, possibly with some mistakes along the way. As more agents end up in the same position the weight of the population average in the utility function increases.

What will the agents' position be in the long run? The stationary distribution of the Markov chain establishes this. For a fixed policy h, configurations with higher social utiliy — the sum of all agents utility — will be realised with higher probability.

Proposition 1 (Stationary Distribution) *The discrete time Markov chain* $\{\boldsymbol{x}_t\}_{n\geq 0}$ *with one step transition probabilities given by* (3.3) *has a unique stationary distribution, which is proportional to the sum of utilities and has the form*

$$\mu[\boldsymbol{x}] = \frac{e^{\beta \sum_\ell \bar{U}_\ell(\boldsymbol{x})}}{\sum\limits_{\boldsymbol{y}\in\Lambda_N} e^{\beta \sum_\ell \bar{U}_\ell(\boldsymbol{y})}}.$$

Looking at the utility in the exponent $\bar{U}_i = hx_i - \frac{\gamma}{2N}\sum_{j\neq i}(x_i - x_j)^2$ it is easy to see that configurations with a high degree of coordination will be more likely. And out of all highly coordinated configurations those that align with the external planner policy h will be favored. This is the same result by Kandori et al. (2008), who also claims that the chain spends most time in these states. This is in fact not the case: while these are the configurations with highest probability, they are only two out of 2^N total possible configuration, and as the next section shows, sometimes the sheer number of lower probability configurations means that these will be more likely to appear.

3.3 Lumping

Lumping consists of reducing the dimensionality of the system by aggregating configurations that share some relevant characteristic, also called an *order parameter*, which in our case is the mean value of agents choice. The following map associates each possible configuration to its mean:

$$m_t := m(\boldsymbol{x}_t) = \frac{\sum_{i=1}^N x_{it}}{N}.$$

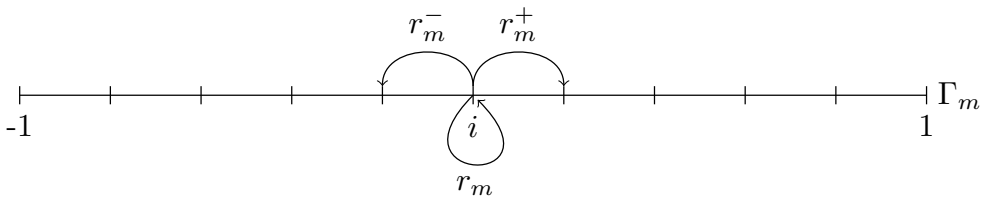


Figure 3.1: Transition rates over the lumped state space Γ_N. These represents the probability that, if the average agents choice is m at time t, it will either increase r_m^+, decrease r_m^- or remain the same r_m, in $t+1$.

Every time one agent revises their choice from x_i to $-x_i$, the mean changes by $\frac{2}{N}$, so m_t takes values in the lumped state space

$$m \in \Gamma_N = \{-1, -1 + 2N^{-1}, \ldots, 1 - 2N^{-1}, 1\},$$

For a given configuration $\boldsymbol{x}$ the share of agents whose choice is currently -1 is $(1 - m(\boldsymbol{x}))/2$. Since all N agents are identical the probability of anyone switching to 1 is given by (3.3). Lumping by summing N times gives the probability that the chain increases:

$$r_m^+ := r(m, m + \frac{2}{N}) = \frac{(1-m)}{2}\left[1 + e^{-2\beta(h+\gamma m + \frac{\gamma}{N})}\right]^{-1} \tag{3.4}$$

and similarly decreases

$$r_m^- := r(m, m - \frac{2}{N}) = \frac{(1+m)}{2}\left[1 + e^{2\beta(h+\gamma m - \frac{\gamma}{N})}\right]^{-1}. \tag{3.5}$$

Given that an agent might not revise his choice there is some residual probability that the mean won't increase, given by

$$r_m = 1 - r_m^- - r_m^+.$$

The new probabilities depend on m alone, therefore Theorem 5 guarantees that $\{m_t\}_{t\geq 0}$ is still Markov with transition matrix

$$R = \begin{bmatrix} & \ddots & \ddots & \ddots & & & \\ \ldots & 0 & r_m^- & r_m & r_m^+ & 0 & \ldots \\ & & & \ddots & \ddots & \ddots & \end{bmatrix}.$$

The stationary distribution is now higly illustrative of the behavior of the chain.

Proposition 2 (Lumped Stationary Distribution) *The lumped chain* $\{m_t\}_{t\geq 0}$

with one step transition probability R has a unique stationary distribution

$$\pi[m] = \frac{e^{-\beta N(-hm-\frac{\gamma}{2}m^2)}}{Z}\left[\binom{N}{\frac{N(1-m)}{2}}2^{-N}\right]. \tag{3.6}$$

It is easier now to interpret the long term behavior of the chain. Highly coordinated configurations, ones where $|m|$ is larger, will have higher weight in the exponent compared to less coordinated ones, but this probability is weighted by the number of configurations that share the same amount of coordination. Simple combinatorics tells us that highly coordinate configurations represent a tiny proportion out of the 2^N available configurations [4]. To understand whether the combinatorial term or the term in the exponent prevail we can write the proportional component of the measure as

$$f_N(m) = -hm - \frac{\gamma}{2}m^2 - \frac{1}{N\beta}\ln\left[\binom{N}{\frac{N(1-m)}{2}}2^{-N}\right],$$

which in the limit[5] of a large population N converges to the potential function:

$$f(m) = -hm - \frac{\gamma}{2}m^2 + \frac{1}{\beta}[(1+m)\ln(1+m) + (1-m)\ln(1-m)], \tag{3.7}$$

The potential function in (3.7) can be interpreted[6] as one over the probability of observing an average position m in the long run. Depending on the relative size of γ and β, when there's no policy, the potential has one of the three shape plotted in Figure 3.2. Using the negative sign convention[7], the measures places more probability on the point where the potential is lower, hence the potential's minima represents the most likely states, as well as those where the chain will spend most of its time.

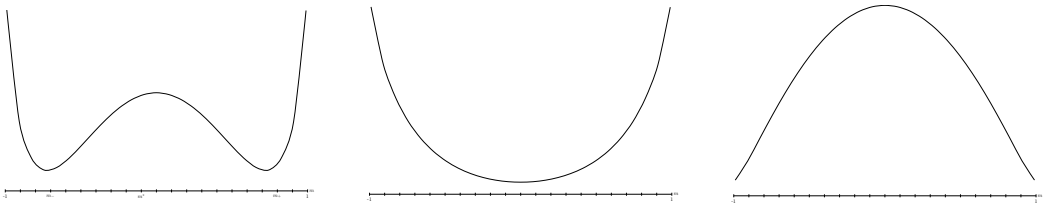
It is useful to understand how the potential function depends on β and γ, leaving the policy $h = 0$ for now. When the coordination motive γ is low, Figure 3.4b, the function has a unique minima: since agents have little interest in coordinating the most likely value of m will be close to zero, that is about half of the population will hold one position. With strong coordination motive, but low ra-

[4]The combinatorial term in the measure precisely counts the number of configurations. The tension between probability accruing to a configurations due to the underlying dynamics that determine agent flips and the number of configuration is what is usually referred as the energy-entropy balance in the statistical mechanics literature.

[5]Note that in the limit the combinatorial term is an even function.

[6]We are employing the negative convention, so the potential function represents one over the probability.

[7]This is a common convention in statistical mechanics, due to the fact that the energy is a negative quantity that is miminized.



(a) Low β and large γ **(b)** Low γ **(c)** Large β

Figure 3.2: The potential function f representing one over the probability of observing a configuration with average $m \in \Gamma_N$. Planner policy is set to $h = 0$. **(a)** Low rationality parameter and strong coordination motives leads to two minima, which imply the existence of two attracting region, with high but not full coordination. **(b)** Very low coordination, $\gamma \approx 0$ entails a unique minima at $m = 0$ where half the agents are adopting one position. **(c)** When agents are strongly rational the states of full coordination become strongly preferred, assuming some minimal amount of coordination γ.

tionality β, Figure 3.4a, then the potential function will have a double well shape. If both parameter are strong than the potential will have two non differentiable minima located at the edge of the state space, Figure 3.2c, and the states with full coordination will be the most likely ones.

The policy also matters in determining the number of minima of $f(m)$. If h is sufficiently small the overall behavior of the function is unchanged, but the function is skewed in the direction of the sign of the policy. When h becomes large enough then the function will display a single minima.

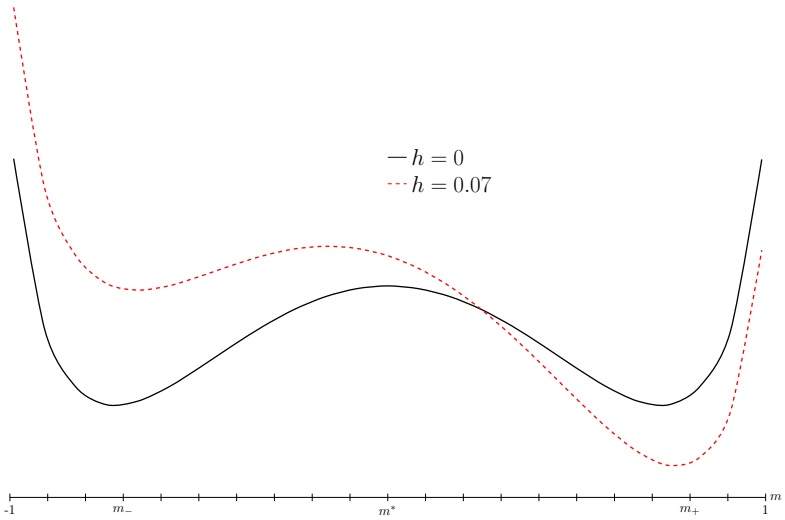
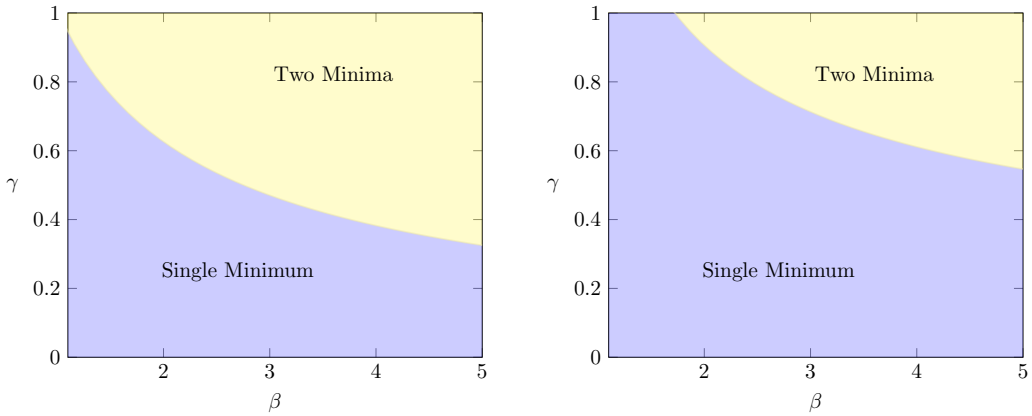


Figure 3.3: Potential figures for two different values of the policy $h = 0$ and $h = 0.07$. The effect of the policy is to skew the distribution of agents average choice. In particular, when two wells are present, one will become less deep as h increases, eventually disappearing.

The range of h of where there are a multiplicity of equilibria is given in the



(a) $h = 0.002$ **(b)** $h = 0.2$

Figure 3.4: Relation between coordination motive γ and rationality parameter β in determining the presence of one or more minima in the potential function (implying one or more long-lived equilibria). A higher policy h reduces the frequency of multiple minima in the state space by shifting the edge between the two regions, according to Proposition 3.

following proposition and pictured in Figure 3.4.

Proposition 3 (Multiple minima) *For $\beta > 1$ when h is within the set*

$$\mathcal{H}(\beta, \gamma) = \mp \left(\gamma \sqrt{1 - \frac{1}{\beta}} - \frac{1}{\beta} \operatorname{atanh} \sqrt{1 - \frac{1}{\beta}} \right) \subset [-1, 1]$$

the potential function f displays multiple minima.

3.4 Long Lived Equilibria

The stationary distribution describes the long term behavior of all agents: given any initial configuration, after a long enough time, the probability of observing a certain average position in the population is given by (3.6). So far we have described the stationary, or long term behavior, of our ensemble of agents and we know that depending on how often they receive preference shocks, the strength of the external policy and how much importance they place on other choices the distribution might be unimodal or bi-modal. Intuitively, when the distribution places a large amount of probability on two configurations it must be the case that the chains also spends more time there. Once the chain is equilibrated and reaches either minima, it is said to have achieved a *long lived equilibrium*. How long does the chain spend in such a position? How long it takes for the chain to abandon the current one and settle into a new one? Using the potential function we can answer these questions.

In general: the expected time to move between two points where the potential is decreasing is polynomial, scaling approximately with the number of agents N. Instead, the expected time to move upward the potential, is exponentially large in the population size.

Let us recall the definition of metastabililty from Section 2.2.4.

Definition 11 (Metastability) *A family of Markov chain indexed by N is called metastable if there exists a collection of disjoint sets $B_i \subset \Omega$, such that*

$$\frac{\sup_{m \notin \cup_i B_i} \mathbb{E}_m[\tau_{\cup_i B_i}]}{\inf_i \inf_{m \in B_i} \mathbb{E}_m[\tau_{\cup_j B_j \setminus B_i}]} = o(1), \ N \to \infty. \tag{2.8}$$

What this means, is that a region of the state space is called metastable when the average time the chain takes to leave the region is much larger than the time it takes to enter it from outside. The $o(1)$ is interpreted to mean that the ratio between the entry time and the exit time from the region will go to zero as $N \to \infty$.

There are many collection of sets that satisfy the definition of metastability. We define long-lived equilibria to be the smallest subsets of the metastable sets when the population size diverges.

Definition 15 (Long-lived equilibria) *A long-lived equilibrium is a the smallest possible collection of sets that exhibits metastability in the limit of large N.*

The next proposition connects minima of the potential function to the metastable sets, showing that the two notions are interchangeable.

Proposition 4 (Minima of the potential identify the long lived equilibria) *consider the Markov chain $\{m_t\}_{t \geq 0}$ over Γ_N with transition matrix r and potential function*

$$f(m) = -hm - \frac{\gamma}{2} m^2 + \frac{1}{\beta}[(1+m)\ln(1+m) + (1-m)\ln(1-m)].$$

If $\mathfrak{M} \subset \Gamma_N$ is a collection of metastable sets for the chain and m^* is a point in Γ_N closest point to $\hat{m} \in \arg\min_{m \in \Gamma_N} f(m)$, then $m^* \in \mathfrak{M}$.

This immediatly implies that the minima of the potential function[8] constitute the long lived equilibria. As a direct consequence of the proof for Proposition 4 we obtain a statement which gives us the order of the expected duration of a long lived equilibria, as well as the time to enter into one.

[8] Or the element of Γ_N which are close to it.

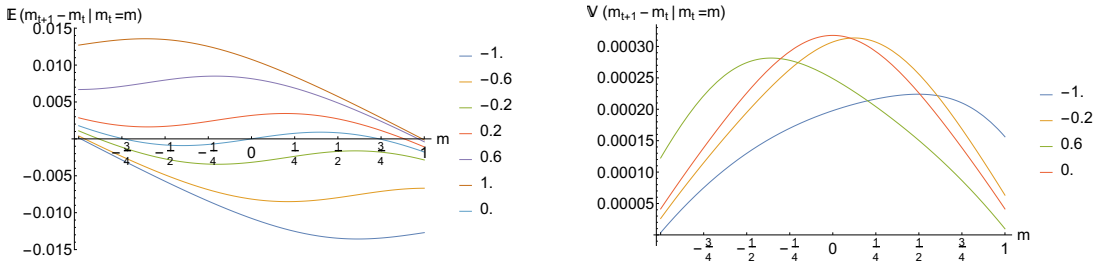
Proposition 5 (Duration of long lived equilibria) *When the chain exhibits more than one long lived equilibria:*

(i) The average time to leave one equilibria is exponential in N.

(ii) The average time to enter one equilibria is polynomial in N.

3.5 Drift

One last useful tool to analyse the system is the the average displacement of m_t after one time step given that the chain started at m



(a) Average displacement **(b)** Variance of displacement

Figure 3.5: (a) Shows the drift d_m, i.e. the average displacement of m_t when the chain starts in state m . Zeros of the drift coincide with minima of the potential function, see for example Fig.3.3. **(b)** Variance of the displacement of m_t when the chain starts at m. Colored lines indicate different values of the policy h.

$$\begin{aligned}\mathbb{E}[m_{t+1} - m_t] &= \left(m + \frac{2}{N}\right) r_m^+ + m r_m + \left(m - \frac{2}{N}\right) r_m^- - m = \\ &= \frac{2}{N}\left(r_m^+ - r_m^-\right)\end{aligned}$$

We denote this quantity as:

$$d_m := \frac{2}{N}\left(r_m^+ - r_m^-\right) \tag{3.8}$$

and call it the average drift. Clearly, the drift depends on h, hence we should sometimes write d_m^h when a specific value of policy is being applied. In particular, when the null policy $h = 0$ is applied, we will write d_m^0.

The drift relates the rates to the stationary measure by behaving as a discrete derivative of the potential function f. This is easy to see, since the zeros of the drift identify the stationary point on f. Set the drift to zero

$$d_m := \frac{2}{N}\left(r_m^+ - r_m^-\right) = 0$$

which can be rewritten as

$$\ln \frac{1+m}{1-m} = -\beta(h + \frac{\gamma}{2}m).$$

This equation is the same as one obtains by taking the derivative of the potential and setting it to zero, and can be reworked to the more familiar (Durlauf, 1996) form

$$m = \tanh\left[\beta\left(h + \frac{\gamma}{2}\right)m\right].$$

3.6 Proofs

Proposition 1 (Stationary Distribution) *The discrete time Markov chain* $\{\boldsymbol{x}_t\}_{n\geq 0}$ *with one step transition probabilities given by* (3.3) *has a unique stationary distribution, which is proportional to the sum of utilities and has the form*

$$\mu[\boldsymbol{x}] = \frac{e^{\beta\sum_\ell \bar{U}_\ell(\boldsymbol{x})}}{\sum\limits_{\boldsymbol{y}\in\Lambda_N} e^{\beta\sum_\ell \bar{U}_\ell(\boldsymbol{y})}}.$$

Proof: If a chain is reversible with respect to a measure then it is also stationary. The reversibility equation (2.5) is trivially satisfied whenever $\boldsymbol{x} = \boldsymbol{y}$ and whenever $|\boldsymbol{x}-\boldsymbol{y}| > 2$, therefore from the reversibility equation we only need to check

$$\frac{P_{\boldsymbol{yx}}}{P_{\boldsymbol{xy}}} = \frac{\mu(\boldsymbol{x})}{\mu(\boldsymbol{y})},$$

for $\boldsymbol{y} = \boldsymbol{x}^i$, indeed using the fact that $y_j = x_j$ for all $j \neq i$ and $y_i = -x_i$

$$\frac{P_{\boldsymbol{yx}}}{P_{\boldsymbol{xy}}} = \frac{e^{\beta V_i(\boldsymbol{x})}}{e^{\beta V_i(\boldsymbol{y})}} = e^{2\beta\left(hx_i + \frac{\gamma}{N}x_i\sum_{j\neq i}x_j\right)}$$

$$
\begin{aligned}
\frac{\mu(\boldsymbol{x})}{\mu(\boldsymbol{y})} &= \exp\beta\left\{\sum_{\ell} h(x_\ell - y_\ell) - \frac{\gamma}{2N}\sum_{\ell}\sum_{j\neq i}\left[(x_\ell - x_j)^2 - (y_\ell - y_j)^2\right]\right\} \\
&= \exp\beta\left\{2hx_i - \frac{\gamma}{2N}\sum_{\ell}\sum_{j\neq i}\left[(x_\ell - x_j)^2 - (y_\ell - y_j)^2\right]\right\} \\
&= \exp\beta\left\{2hx_i - \frac{\gamma}{2N}\left[\sum_{j\neq i}(x_i - x_j)^2 - (y_i - y_j)^2\right] - \sum_{\ell\neq i}\sum_{j\neq i}\left[(x_\ell - x_j)^2 - (y_\ell - y_j)^2\right]\right\} \\
&= \exp\beta\left\{2hx_i - \frac{\gamma}{2N}\sum_{j\neq i}\left[(x_i - x_j)^2 - (y_i - y_j)^2\right]\right\} \\
&= \exp 2\beta\left\{hx_i + \frac{\gamma}{N}x_i\sum_{j\neq i}x_j\right\}
\end{aligned}
$$

□.

Proposition 2 (Lumped Stationary Distribution) *The lumped chain* $\{m_t\}_{t\geq 0}$ *with one step transition probability* R *has a unique stationary distribution*

$$\pi[m] = \frac{e^{-\beta N(-hm-\frac{\gamma}{2}m^2)}}{Z}\left[\binom{N}{\frac{N(1-m)}{2}}2^{-N}\right]. \tag{3.6}$$

Proof: We construct the stationary distribution of the chain by solving the reversibility equation recursively. It is convenient to index values of m as

$$m_i = -1 + \frac{2}{N}i,\ i \in 0, \ldots, N$$

The reversibility equation can be written as

$$\pi[m_i] = \frac{R_{m_{i-1},m_i}}{R_{m_i,m_{i-1}}}\pi[m_{i-1}] = \frac{r^+_{m_{i-1}}}{r^-_{m_i}}\pi[m_{i-1}]$$

substituting recursively we get

$$\pi[m_i] = \prod_{\ell=1}^{i}\frac{r^+_{m_{\ell-1}}}{r^-_{m_\ell}}\pi[m_0].$$

This gives an expression for the measure $\pi[m_i]$ in terms of some reference[9] $\pi[m_0]$.

[9] Any other index could be used, the zeroth one is picked for convenience.

Using the normalization $\sum_i \bar{\pi}[m_i] = 1$ we obtain

$$\pi[m_0] = \left[\sum_{j=0}^{N} \prod_{\ell=1}^{i} \frac{r^+_{m_{\ell-1}}}{r^-_{m_\ell}} \right]^{-1}$$

The transition probabilities can be written as

$$r^\pm_m = \frac{(1 \mp m)}{2} \frac{e^{\pm\beta(h_m + \gamma m \pm \gamma/N)}}{\cosh(\beta(h_m + \gamma m \pm \gamma/N))}$$

so that

$$\begin{aligned}
\prod_{\ell=1}^{i} \frac{r^+_{m_{\ell-1}}}{r^-_{m_\ell}} &= \prod_{\ell=1}^{i} \left(\frac{1 - m_{\ell-1}}{1 + m_\ell} \right) \frac{e^{-\beta(h+\gamma m_{\ell-1}+\gamma/N)}}{e^{\beta(h+\gamma m_\ell - \gamma/N)}} \frac{\cosh(\beta(h + \gamma m_\ell \quad - \gamma/N))}{\cosh(\beta(h + \gamma m_{\ell-1} + \gamma/N))} \\
&= e^{\beta \sum_{\ell=1}^{i} (2h + \gamma m_\ell + \gamma m_{\ell-1})} \prod_{\ell=1}^{i} \left(\frac{1 - m_{\ell-1}}{1 + m_\ell} \right) \frac{\cosh(\beta(h + \gamma m_\ell - \gamma/N))}{\cosh(\beta(h + \gamma m_\ell - \gamma/N))} \\
&= e^{\beta \sum_{\ell=1}^{i} 2[(h+\gamma m_\ell) - \frac{2\gamma}{N}]} \prod_{\ell=1}^{i} I_\ell = e^{-\beta[2(hi + \gamma(-i + \frac{i(i+1)}{N})) - \frac{2\gamma}{N} i]} \prod_{\ell=1}^{i} I_\ell \\
&= e^{\beta[2(hi + \gamma(-i + \frac{i^2}{N}))]} \prod_{\ell=1}^{i} I_\ell = e^{-\beta N[(hm_i + \frac{\gamma}{2} m_i^2] + \beta N(-h + \frac{\gamma}{2})} \prod_{\ell=1}^{i} I_\ell \\
&= e^{\beta N(hm_i + \frac{\gamma}{2} m_i^2)} \prod_{\ell=1}^{i} I_\ell
\end{aligned}$$

where the last two equalities follow from adding and subtracting appropriate quantities and by the fact that we can discard all elements that do not depend on i since they can be collected in the denominator. The combinatorial term in (3.6) weights the measure by the number of possible way in which state m can be achieved. First note that the combinatorial terms, using the definition of m_i, rewrites as

$$\binom{N}{N\left(\frac{1-m_\ell}{2}\right)} = \binom{N}{N-i} = \frac{N!}{(N-i)!(N-N+i)!} = \frac{N!}{i!(N-i)!} = \binom{N}{i},$$

which is equal to $\prod_{l=1}^{i} I_\ell$, indeed

$$\prod_{l=1}^{i} I_\ell = \prod_{\ell=1}^{i} \left(\frac{1 - m_{\ell-1}}{1 + m_\ell} \right) = \prod_{\ell=1}^{i} \left(\frac{N - \ell + 1}{\ell} \right) \left(\frac{(N-i)!}{(N-i)!} \right) = \binom{N}{i} \square.$$

Proposition 3 (Multiple minima) *For $\beta > 1$ when h is within the set*

$$\mathcal{H}(\beta, \gamma) = \mp \left(\gamma \sqrt{1 - \frac{1}{\beta}} - \frac{1}{\beta} \operatorname{atanh} \sqrt{1 - \frac{1}{\beta}} \right) \subset [-1, 1]$$

the potential function f displays multiple minima.

Proof: The range is obtained by imposing that h is such that the potential function f in the limit of large N exhibits two wells. Letting the first derivative of f with respect to m be zero

$$m = \tanh[\beta(\gamma m + h)] \tag{3.9}$$

f has multiple minima when this equation has multiple solutions. Changes in h makes two stationary points close together, until eventually they merge into one before disappearing. This happens when (3.9) is satisfied and the slope of its right hand matches the slope of the left hand side, that is when

$$\frac{\partial}{\partial m} \tanh[\beta(\gamma m + h)] = 1. \tag{3.10}$$

The solution of the system of equations

$$\begin{cases} m = \tanh[\beta(\gamma m + h)] \\ 1 = \beta \left[1 - \tanh^2[\beta(\gamma m + h)] \right] \end{cases},$$

yields the interval above. □

Proposition 4 (Minima of the potential identify the long lived equilibria) *consider the Markov chain $\{m_t\}_{t \geq 0}$ over Γ_N with transition matrix r and potential function*

$$f(m) = -hm - \frac{\gamma}{2} m^2 + \frac{1}{\beta} [(1 + m) \ln(1 + m) + (1 - m) \ln(1 - m)].$$

If $\mathfrak{M} \subset \Gamma_N$ is a collection of metastable sets for the chain and m^* is a point in Γ_N closest point to $\hat{m} \in \arg\min_{m \in \Gamma_N} f(m)$, then $m^* \in \mathfrak{M}$.

Proof: Given some collection of metastable set B, we want to prove that whenever we start the chain from as close as possible to the minima of $f(m)$ the average time to leave B is diverging faster than the time it will take to enter B from outside, so that the ratio in the definition of metastability (2.8) goes to zero in the limit of large N.

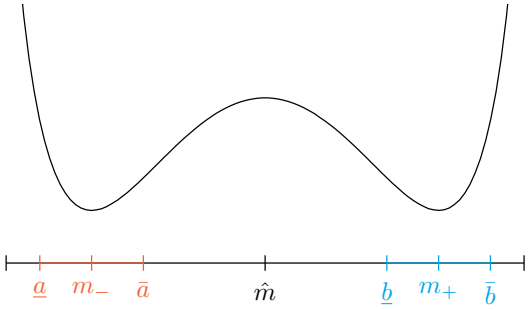


Figure 3.6: Metastable sets B_1 and B_2

Assume that $h \in \mathcal{H}(\beta, \gamma)$, according to proposition 3 the potential function has two minima: m_+ and m_-. Also assume that there's a metastable collection $B \subset \Gamma_N$ formed by two metastable interval around the minima: $A = [\underline{a}, \overline{a}]$ and $B = [\underline{b}, \overline{b}]$. For simplicity we can let $h = 0$ without loss of generality.

First, let's show that the denominator in the definition of metastability (2.8) is diverging exponentially in N. Recalling the formula to compute expected hitting time and without loss of generality, consider the average time to leave the metastable set B starting from m_+

$$\mathbb{E}_{m_+}[\tau_{\underline{b}}] = \sum_{\substack{m,m' \in \Gamma \\ m \leq m' \\ \underline{b} < m \leq m_+}} \frac{\mu[m']}{\mu[m]} \frac{1}{r_m^-}, \ \underline{b} < m_+. \tag{3.11}$$

The ratio of measures in (3.11) takes the form

$$\frac{\mu[m']}{\mu[m]} = e^{\beta N[f_N(m) - f_N(m')]}.$$

The exponential term is maximal at $m = \underline{b}$ and $m' = m_+$. Collecting the maximum, equation (3.11) becomes

$$\begin{aligned} &= e^{\beta N[f_N(\underline{b}) - f_N(m_+)]} \frac{1}{r_{\underline{b}}^-}[1 + o(1)] \\ &\qquad \times \sum_{\substack{|m - \underline{b}| < \epsilon \\ |m' - m_+| < \epsilon}} e^{\beta N[f_N(m) - f_N(\underline{b})] - \beta N[f_N(m') - f_N(m_+)]} \end{aligned} \tag{3.12}$$

where we use the fact that r_m^- doesn't depend on N, is strictly positive and bounded below one. It can be shown[10] that $f_N \to f$, approximating f_N with f the error once

[10] Using Stirling's formula.

raised to the exponent is given by

$$e^{f_N(m)-f(m)} = [1+o(1)]\sqrt{\frac{\pi N(1-m^2)}{2}}$$

and therefore

$$\begin{aligned} &= e^{\beta N[f(\underline{b})-f(m_+)]}\frac{1}{r_{\underline{b}}^-}[1+o(1)]\sqrt{\frac{(1-m_+^2)}{(1-\underline{b}^2)}} \\ &\qquad \times \sum_{\substack{|m-\underline{b}|<\epsilon \\ |m'-m_+|<\epsilon}} e^{\beta N[f(m)-f(\underline{b})]-\beta N[f(m')-f(m_+)]} \end{aligned} \tag{3.13}$$

Taylor expanding the terms at the exponent:

$$f(m)-f(\underline{b}) = f'(\underline{b})(m-\underline{b}) + \frac{1}{2}f''(\underline{b})(m-\underline{b})^2 + O((m-\underline{b})^3)$$

$$f(m')-f(m_+) = f'(m_+)(m'-m_+) + \frac{1}{2}f''(m_+)(m'-m_+)^2 + O((m'-m_+)^3)$$

using the fact that $f'(m_+)=0$ the summation now becomes:

$$\begin{aligned} &= e^{\beta N[f(\underline{b})-f(m_+)]}\frac{1}{r_{\underline{b}}^-}[1+o(1)]\sqrt{\frac{(1-m_+^2)}{(1-\underline{b}^2)}} \\ &\qquad \times \sum_{\substack{|m-\underline{b}|<\epsilon \\ |m'-m_+|<\epsilon}} e^{\beta N[f'(\underline{b})(m-\underline{b})+\frac{1}{2}f''(\underline{b})(m-\underline{b})^2-\frac{1}{2}f''(m_+)(m-m_+)^2]}. \end{aligned}$$

Using the following substitution of variable

$$u = \sqrt{N}(m-\tilde{z}),\ u' = \sqrt{N}(m-m_+)$$

we can turn the summations into integrals, given our sampling in each integral is $\sqrt{N}/2$

$$\frac{N}{4}\int\int e^{\sqrt{N}Au+Bu^2-Cu'^2}dudu'.$$

Both A and B are negative, while C is stricly positive. The integral therefore converges to some finite quantity, showing that the expected time to leave a metastable set is

$$\mathbb{E}_{m_+}[\tau_{\underline{b}}] = O(e^{\beta N}N). \tag{3.14}$$

It remains to show that the numerator of definition (2.8), the expected time of entering the metastable set, is always polynomial in N in the presence of multiple

minima. Take some $z \notin B$ as the starting point and without loss of generality compute the time to enter B_1 as

$$\mathbb{E}_z[\tau_{\overline{a}}] = \sum_{\substack{m,m' \in \Gamma \\ m \leq m' \\ \overline{a} \leq m \leq z}} e^{\beta N[f_N(m) - f_N(m')]} \frac{1}{r_m^-}$$

this term is maximal when $m = m' = z$, that is the largest element in the summation above is equal to 1. A similar computation as the one for the denominator shows that the summation converges to a finite value with an N prefactor, showing that

$$\mathbb{E}_z[\tau_{\overline{a}}] \approx O(N).$$

□

Proposition 5 (Duration of long lived equilibria) *When the chain exhibits more than one long lived equilibria:*

(i) The average time to leave one equilibria is exponential in N.

(ii) The average time to enter one equilibria is polynomial in N.

Proof: Follows directly from the proof of the previous proposition □.

Chapter 4

Planner Problem with Lagged Sampling

In this chapter the policy h can be set by an external planner having a preference $\eta \in \{-1, 1\}$ for one of the two positions. The planner pays a cost proportional to the squared[1] distance of the average agent position from its preference, as well as a squared cost for the policy it decides to set. The per-period cost function, assuming in t the lumped chain is at $m_t = m$, is given by:

$$C(m, h) = \frac{c}{2}h^2 + (m - \eta)^2.$$

The planner then wishes to minimize the average discounted cost over the infinite horizon which constitutes the lifetime of the chain. In terms of stochastic optimization the problem can be expressed with the value function[2] which associates the average cost under the optimal policy given each possible initial condition[3]

$$\begin{aligned} V_m :&= \inf_{\bar{h}} \mathbb{E}^{\bar{h}} \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \,\middle|\, m_0 = m \right] = \\ &= \inf_{\bar{h}} \mathbb{E}_m^{\bar{h}} \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \right], \end{aligned} \tag{4.1}$$

[1] Squared costs are used for convenience as they yield a function which is differentiable and convex, the latter being the typical requirement associated to a cost function.

[2] Despite the name the value function is used to denote both reward maximization and cost minimization problem.

[3] Or in other terms, it expresses the optimal average cost, V_m for a planner that has to pick a new policy when the current state of the process is m.

where $\delta \in (0,1)$ is the discount factor, and the notation $\mathbb{E}_m[\cdot] = \mathbb{E}[\cdot|m_0 = m]$ denotes the expectation of the chain m_t with initial condition $m_0 = m$.

In this chapter it is assumed that between each time step the process has equilibrated, implying that the probability of a specific realization of m_t is given by the stationary distribution $\pi[m]$. Equivalently, we are saying that the planner can only "sample" the population average and review its policy[4] with a considerable delay. The consequence of this assumption, which will be relaxed in Chapter 5, is that the optimal policy is a single value $\bar{h}_m = \bar{h}$ independent of the location of the process.

4.1 Optimality equation

It is easy to see that the per-period cost function is bounded, since $|m| \leq 1$ and $h \in [-1,1]$ by assumption. Therefore, from Chapter 2.3, the optimal policy is unique and the value function is equivalent to the optimality equation:

$$V_m = \min_{h\in[-1,1]} \left\{ C(m,h) + \delta \sum_{n\in\Gamma_N} \pi_n(h) V_n \right\}. \tag{4.2}$$

Under the optimal policy $\bar{h}$ the value function can be written in matrix notation

$$V = C + \delta \Pi V$$

where Π is the matrix with the stationary distribution on the diagonal

$$\Pi = \begin{bmatrix} \ddots & & \\ & \pi_m & \\ & & \ddots \end{bmatrix},$$

and zero everywhere else. This can be solved algebraically to obtain

$$V = (I - \delta\Pi)^{-1} C$$

so the component-wise solution of the value function is given by

$$V_m = \frac{\bar{h}^2 + (m-\eta)^2}{1 - \delta\pi_m}. \tag{4.3}$$

[4] Any cost sustained inbetween sampling periods is a fixed cost that cannot be reduced and therefore does not affect the minimization.

Hence, the value of the problem at m is the discounted cost of being at m, including the cost of the optimal policy to be supplied, weighted by the probability of ending there. The weighted average interpretation becomes clear by rewriting the denominator of Eq. (4.3) as a geometric sum

$$V_m = \sum_{t=0}^{\infty} (\delta\pi_m)^t [\bar{h}^2 + (m-\eta)^2]$$

which is also useful to establish the following proposition.

Proposition 6 (Value function is convex in the policy) *The value function V_m is convex with respect to the policy $\bar{h}$.*

Proof: Rewrite equation (4.3) as

$$V_m(h) = \sum_{k=0}^{\infty} (\delta\pi)^k [h^2 + (m-\eta)^2]$$

where we omit dependencies of $\pi := \pi_m(h)$ for readability, then

$$\begin{aligned}\frac{\partial^2 V_m(h)}{\partial^2 h} = \sum_{k=0}^{\infty} \Big\{ & k(h^2 + (m-\eta)^2) \left[(k-1)(\delta\pi)^{k-2} (\delta\pi')^2 + (\delta\pi)^{k-1} \delta\pi'' \right] \\ & + 4hk(\delta\pi)^{k-1}\delta\pi' + (\delta\pi)^k \Big\} \geq 0.\end{aligned}$$

This inequality holds since the leading term is positive and diverges as k^2. □

This property of the curvature of the value function critically depends on the fact that the optimal policy is a constant value for all m, which in turns depend on the maintained assumption that the chain has reached equilibration before the planner can sample a new realization.

4.2 Characterizing the optimal policy

Convexity of the value function guarantees that the solution of the first order condition is a unique minimum, hence the optimal policy can be obtained by finding a first order condition for V.

Proposition 7 (First order condition) *The optimal policy is given by*

$$\bar{h} = \min[\max[\tilde{h}, -1], 1]$$

where $\tilde{h}$ is the solution of the implicit equation

$$h = -\frac{\delta\beta N}{c}\left\{\mathbb{S}k^h[m_t] - 2\eta\mathbb{V}^h[m_t]^{-1/2}\right\}\mathbb{V}^h[m_t]^{3/2}, \tag{4.4}$$

which is the first order condition for the value function.

Proof: For a generic time index t, under our maintained assumption m_t is distributed according to the stationary measure

$$\pi_{m_t}(h) = \frac{e^{-\beta N f(m_t,h)}}{Z}$$

where $\frac{\partial}{\partial h}f = m_t$ and Z is the appropriate normalization factor. If h were left free in variation, then $V_{m_t}(h)$ would be a convex function of h according to Proposition 6 hence setting its derivative with respect to h to 0 identifies the unique minima.

$$\begin{aligned}
\frac{\partial}{\partial h}\tilde{V}_m &= 2h + \delta\sum_{m_t\in\Gamma_N}\left[\beta N\frac{e^{-\beta N f(m_t)}}{Z}\left(m_t - \sum_{\ell}\ell_t\frac{e^{-\beta N f(\ell_t)}}{Z}\right)\right]V_{m_t}\\
&= 2h + \delta\beta N\sum_{m_t\in\Gamma_N}\pi_{m_t}(m_t - \mathbb{E}[m_t])V_{m_t}\\
&= 2h + \delta\beta N\mathbb{C}ov[m_t, V_{m_t}]\\
&= 2h + \delta\beta N\mathbb{C}ov[m_t, \bar{h} + (m_t-\eta)^2 + \mathbb{E}[V]]\\
&= 2h + \delta\beta N\mathbb{C}ov[m_t, (m_t-\eta)^2].
\end{aligned}$$

The last equality is obtained because the optimal policy $\bar{h}$ within V_{m_t} is independent of the realization of m_t. Expanding the covariance term and collecting $\mathbb{V}[m_t]^{\frac{3}{2}}$ yields the first order condition. □

Hence, the first order condition, and therefore the optimal policy depend on the skewness of the distribution π as well as on its variance. Both these moments depend on h itself, and due to their exponential form there can be no explicit solution to this equation. Further the presence of the population size N means that whenever the population is large, given that the moments are finite for finite parameters, the optimal policy will be picked at the extreme of the admissible range as long as cost are constant in either β or N. In turn, these means that for sufficiently low cost the policy will always be strong enough to remove one of the two attracting points and as a result only the equilibrium closer to the target of the planner survives.

When the marginal costs of a unit of policy are proportional to the population size then the optimal policy might not be strong enough and multiple equilibria might survive. Figure 4.1 shows the policy when $c = N$: for low values of the

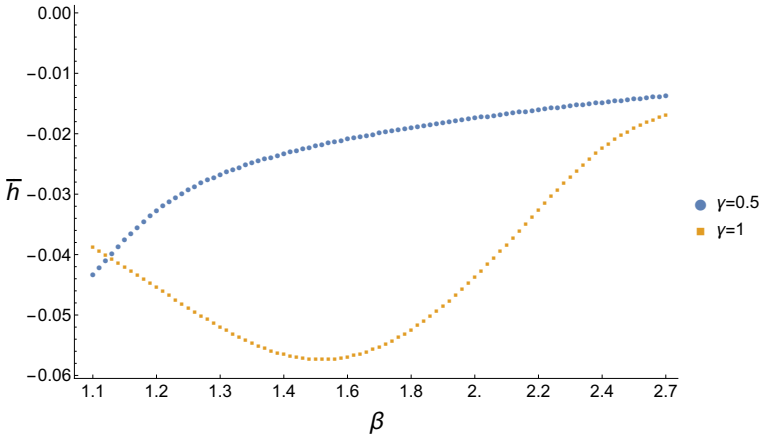


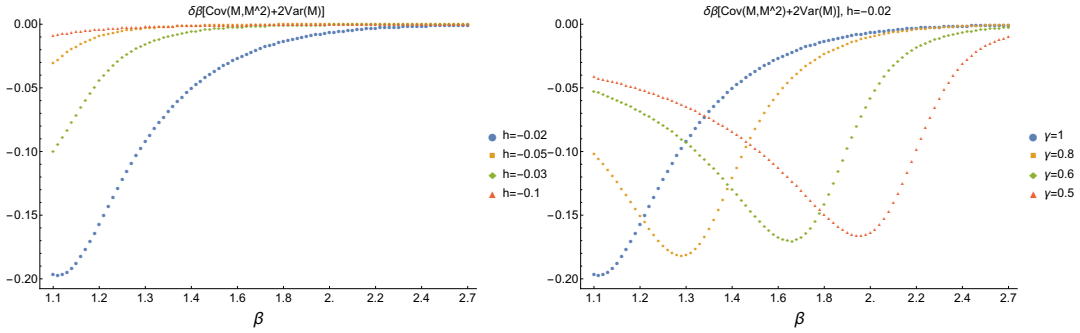
Figure 4.1: Optimal policy $\bar{h}$ when policy marginal costs are proportional to the population size $c = N$ at different level of rationality β and for two different values of the coordination motive γ. Higher policy values are optimal for more rational agents when there is little coordinations among agents ($\gamma = 0.5$). With strong coordination ($\gamma = 1$) the planner benefits from peer pressure effects when agents are not too rational and commit mistakes often.

coordination the optimal policy is monotonic in β, which makes intuitive sense since for low enough values of γ there's a unique long lived equilibrium (see Fig. 3.4b). Instead, when coordination is sufficiently strong the policy becomes non monotonic in β. Past a certain value agents tend to stick more to their current position and additional unit of policy becomes less effective in shifting agents position.

Proposition 8 (Multiple long lived equilibria under the optimal policy) *For a sufficiently large population N and under the optimal policy there exists constants $\nu > \zeta$ such*

- *if $c \geq \delta\beta e^{\nu\beta N} O(N^3)$ there is a multiplicity of long lived equilibria.*
- *if $c \leq \delta\beta e^{\zeta\beta N} O(N^2)$ there is a unique long lived equilibrium and the optimal policy is always either -1 or 1.*

Proof: We would like to obtain the exact behaviour of the right hand side of the first order condition (4.4) as a function of β, which would provide a full analytical description of the optimal policy. The FOC solution, and hence the optimal policy behavior, is plotted numerically in Fig. (4.2b) and (4.2a). Since these cannot be obtained explicitly, we resort to compute upper and lower bound for moments of $M := m_t$ for a generic t.



(a) $\gamma = 1$ **(b)** $h = -0.02$

Figure 4.2: Plot of the right hand side of the first order condition, given by (4.4), that the optimal policy must satisfy. Solutions of $\delta\beta\mathbb{C}ov[M, M^2] - 2\eta\mathbb{V}[M]$ therefore give the behavior of the optimal policy and are plotted below for marginal costs equal to the population size $c = N$ and different values of the coordination motive γ and rationality parameter β.

The upper bound for the $k-th$ order moment can be written as

$$\mathbb{E}[M^k] = \sum_m m^k \frac{e^{-\beta N f(m)}}{Z_\beta} \leq N \max_m m^k \frac{e^{-\beta N f(m)}}{Z_\beta} = N\Omega_k \frac{e^{\beta N \bar{f}_k}}{Z_\beta}$$

and lower bound

$$\mathbb{E}[M^k] = \sum_m m^k \frac{e^{-\beta N f(m)}}{Z_\beta} \geq N \min_m m^k e^{-\beta N f(m)} = N\omega_k \frac{e^{\beta N \underline{f}_k}}{Z_\beta}.$$

Express the right hand side of (4.4) in terms of the covariance and replace $\eta = -1$

$$\begin{aligned}
\mathbb{C}ov[M, (M-\eta)^2] &= \mathbb{C}ov[M, M^2] - 2\eta\mathbb{V}[M] \\
&= \mathbb{E}[M^3] - \mathbb{E}[M]\mathbb{E}[M^2] + 2(\mathbb{E}[M^2] - \mathbb{E}[M]^2) \\
&\leq \frac{1}{Z_\beta}\Big[N\Omega_3 e^{\beta N \bar{f}_3} - (N\omega_1 e^{\beta N \underline{f}_1})(N\omega_2 e^{\beta N \underline{f}_2}) \\
&\qquad + 2(N\Omega_2 e^{\beta N \bar{f}_2} - (N\omega_1 e^{\beta N \underline{f}_1})^2\Big] \\
&= \frac{N}{Z_\beta}\left[\Omega_3 e^{\beta N \bar{f}_3} - N\omega_1\omega_2 e^{\beta N \underline{f}_1 + \underline{f}_2} + 2(\Omega_2 e^{\beta N \bar{f}_2} - N\omega_1^2 e^{2\beta N \underline{f}_1})\right] \\
&\leq \max_{x \in \underline{f}_1, \underline{f}_2, \bar{f}_2, \bar{f}_3} \frac{N}{Z_\beta}\left[\Omega_3 e^{\beta N x} - N\omega_1\omega_2 e^{2\beta N x} + 2(\Omega_2 e^{\beta N x} - N\omega_1^2 e^{2\beta N x})\right] \\
&= N\frac{e^{\nu\beta N}}{Z_\beta}\left[\Omega_3 - N\omega_1\omega_2 + 2(\Omega_2 - N\omega_1^2)\right] \\
&= N^2\frac{e^{\nu\beta N}}{Z_\beta}\left[\frac{\tilde{\Omega}}{N} - \tilde{\omega}\right] = \frac{e^{\nu\beta N}}{Z_\beta}O(N^2)
\end{aligned}$$

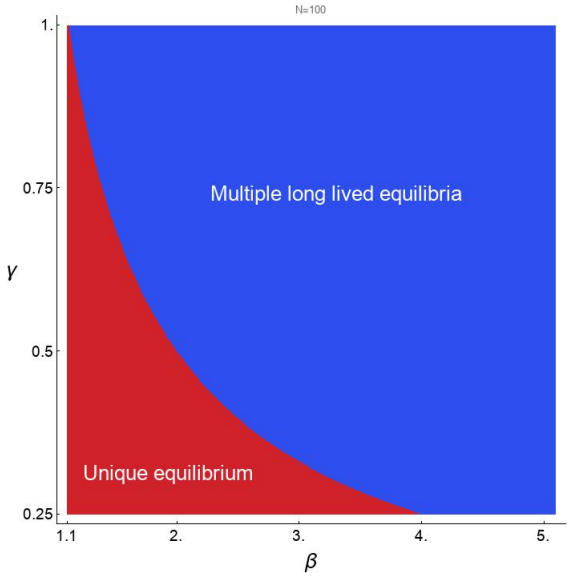


Figure 4.3: Space of the rationality paramater β and coordination motive γ where long lived equilibria persist under the optimal policy $\bar{h}$. This has been numerically obtained for a population of size $N = 80$ with marginal policy costs $c = N$ and discount factor $\delta = 0.9$.

This slack upper bound guarantees that we can control the terms in bracket in equation (4.4) to ensure that the the optimal policy is small enough so that $\bar{h} \in \mathcal{M}(\gamma, \beta)$ which ensures the presence of a multiplicity of long lived equilibria. In a similar way we can obtain the lower bound:

$$
\begin{aligned}
\mathbb{C}ov[M, (M-\eta)^2] &= \mathbb{C}ov[M, M^2] - 2\eta\mathbb{V}[M] \\
&= \mathbb{E}[M^3] - \mathbb{E}[M]\mathbb{E}[M^2] + 2(\mathbb{E}[M^2] - \mathbb{E}[M]^2) \\
&\geq \frac{1}{Z_\beta}\left[N\omega_3 e^{\beta N \underline{f}_3} - (N\Omega_1 e^{\beta N \bar{f}_1})(N\Omega_2 e^{\beta N \bar{f}_2}) + 2(N\omega_2 e^{\beta N \underline{f}_2} - (N\Omega_1 e^{\beta N \bar{f}_1})^2\right] \\
&= \frac{N}{Z_\beta}\left[\omega_3 e^{\beta N \underline{f}_3} - N\Omega_1\Omega_2 e^{\beta N \bar{f}_1 + \bar{f}_2} + 2(\omega_2 e^{\beta N \underline{f}_2} - N\Omega_1^2 e^{2\beta N \bar{f}_1})\right] \\
&\geq \min_{x \in \bar{f}_1, \bar{f}_2, \underline{f}_2, \underline{f}_3} \frac{N}{Z_\beta}\left[\omega_3 e^{\beta N x} - N\Omega_1\Omega_2 e^{2\beta N x} + 2(\omega_2 e^{\beta N x} - N\Omega_1^2 e^{2\beta N x})\right] \\
&= N\frac{e^{\zeta\beta N}}{Z_\beta}\left[\omega_3 - N\Omega_1\Omega_2 + 2(\omega_2 - N\Omega_1^2)\right] \\
&= N^2\frac{e^{\zeta\beta N}}{Z_\beta}\left[\frac{\tilde{\omega}}{N} - \tilde{\Omega}\right] = \frac{e^{\zeta\beta N}}{Z_\beta}O(N^2).
\end{aligned}
$$

□

The upper bound guarantees that the solution of the first order condition is small enough, since costs are higher, for $\bar{h}$ to be in the set $\mathcal{M}(\gamma, \beta)$. Hence,

when costs c are at or above this bounds then the optimal policy is always to small to remove the less preferred long lived equilibria. We know that the presence of adverse long lived equilibria, once the agents choice is within the basin of attraction described by the valley of the potential function f (see Proposition 4), might be extremely costly as the time for the other basin to be reached is exponential in the number of agents. This suggests that in real world situations where herding is prominent the relevant policy advice would be to invest in lowering the cost. Reaching below the lower bound above would guarantee that the optimal policy is always maximal and therefore no long lived equilibria survives, though the bounds presented here are very slack. Indeed, it is sufficient for the costs to be proportional to the population size, to obtain that a large portion of the $\beta \times \gamma$ parameter space only present a single equilibria under the optimal policy.

Chapter 5

Planner Problem with Frequent Sampling

Up until now the policy was allowed to change only after a fixed lag ensuring that the underlying Markov chain describing the agents behavior had reached its stationary distribution. In this chapter the policy is allowed to change at each time step t. A planner is in charge of setting its value every time step after having observed the current value of the process m_t. After the planner sets the new policy $h_{t+1} \in [-1, 1]$ a new value m_{t+1} is realized. The most important consequences of the new assumption is that the optimal policy will now be a function of the state.

The value function is now an average of the value of near-neighbours states on the lattice Γ_N weighted by the transition probability R of the lumped process and is no longer convex in the policy parameter. In general, properties of the value function become much harder to compute analitically.

In this chapter, I compute the first order condition under the new assumption and prove by coupling that the value function is increasing over the lumped lattice. These are then used to show that the optimal policy is non-monotonic over Γ_N and in particular that the highest optimal policy value is always applied after more than half of the population shares the position opposite to planner target state. Lastly, the question of when long lived equilibria survive the application of the optimal policy is numerically answered. The per-period cost function remains the same as in the previous chapter: the sum of the squared distance of the current state from the planner's target state η plus a marginal cost $c/2$ for every unit of squared policy. Hence the per-period cost function:

$$C(m, h) = \frac{c}{2}h^2 + (m - \eta)^2.$$

The planner then wishes to minimizes the average discounted cost over the infinite horizon which constitutes the lifetime of the chain. The problem can be expressed as a value function which associates the average cost under the optimal policy given each possible initial condition of the process:

$$\begin{aligned} V_m :&= \inf_h \mathbb{E}^h \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \,\middle|\, m_0 = m \right] = \\ &= \inf_h \mathbb{E}^h_m \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \right], \end{aligned} \tag{5.1}$$

where we use the notation $\mathbb{E}_m[\cdot] = \mathbb{E}[\cdot|m_0 = m]$ to denote the expectation over the chain m_t with initial condition $m_0 = m$. In principle the infimum of (5.1) is taken over all possible collection of infinite sequences of control, each sequence being the sequence of controls applied on one of the (uncountably many) path the chain m_t can take. In practice the optimal policy at time t will depend exclusively on the state of the chain at that point, therefore we will be looking for a function $\bar{h} : \Gamma_m \to [-1, 1]$ from the state space to the space of possible controls. Applying this policy induces the stochastic sequence of controls $\bar{h}_t = \bar{h}(m_t)$, yielding the Markov chain with transition probabilities $R(\bar{h}(m))$ which solves the planner problem.

5.1 Computing the Optimal Policy

The minimal payoff is equivalent, by Theorem 2.14, to the optimality equation

$$V_m = \min_h \left\{ C(m, h) + \delta \sum_{n \in \Gamma_N} R_{mn}(h) V_n \right\}.$$

Using the definition of the rates of the lumped process in Eq.(3.4), this can be written as

$$\begin{aligned} V_m &= \min_h \left\{ C(m, h) + \delta [r^+_m V_{m+2/N} + r^-_m V_{m-2/N} + (1 - r^+_m - r^-_m) V_m] \right\} \\ &= \min_h \left\{ C(m, h) + \delta [r^+_m \Delta V_{m+1} + r^-_m \Delta V_m + V_m \right\}, \end{aligned} \tag{5.2}$$

where $\Delta V_{m+1} = V_{m+2/N} - V_m$ and $\Delta V_m = V_m - V_{m-2/N}$ and the rates r^+_m, r^-_m are the probabilities of the average choice in the population moving up or down, both functions of h defined in Eq.(3.4) and (3.5).

Taking the partial derivative with respect to h of the r.h.s. of Eq. (5.2) we obtain the first order condition that characterizes the optimal policy. Notice that

the recursive definition of V_m does not depend on h.

Proposition 9 (First order condition) *The optimal policy is given by*

$$\bar{h} = \min[\max[\tilde{h}, -1], 1]$$

where $\tilde{h}$ *must satisfy*

$$h = -\frac{\delta}{c}\left[\partial_h r_m^+ \Delta V_{m+1} - \partial_h r_m^- \Delta V_m\right], \tag{5.3}$$

the first order condition for the value function.

The solution of the first order equation need not be unique, indeed the value function might display multiple minima

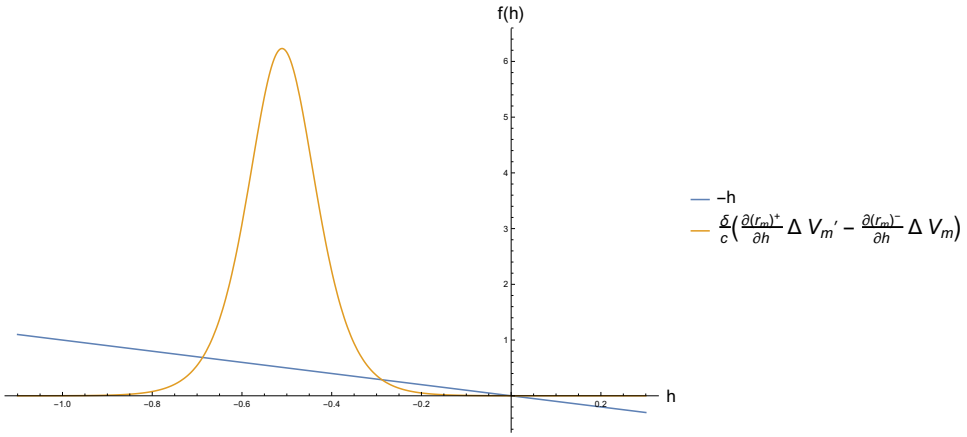


Figure 5.1: Plot of the first order condition showing that the solution to Eq. (5.3) need not be unique.

We obtain the value function numerically employing the policy improvement algorithm described in Section 2.3.5.

We start by writing out the optimality equation (5.1) replacing the minimum with an arbitrary policy $\underline{\pi}$, hence in matrix notation, we have the system

$$V_{\underline{\pi}} = C_{\underline{\pi}} + \delta R_{\underline{\pi}} V_{\underline{\pi}}$$

which can always be solved by matrix inversion since the transition matrix R is stochastic and the discount factor δ is strictly smaller than one

$$V_{\underline{\pi}} = (I - \delta R_{\underline{\pi}})^{-1} C_{\underline{\pi}}.$$

After computing $V_{\underline{\pi}}$ we propose a new policy $\underline{\omega}$ picked so that

$$\underline{\omega} = \arg\min_{\underline{h}} V_{\underline{h}} = \arg\min_{\underline{h}} C_{\underline{h}} + \delta R_{\underline{h}} V_{\underline{\pi}}.$$

Finally, we set $\underline{\pi} = \underline{\omega}$ and repeat until a desired level of tolerance is achieved. Once $V_{\underline{\omega}} = V_{\underline{\pi}}$ then by Theorem 11, $\underline{\omega}$ is the optimal policy.

5.2 Characterization of the optimal policy

There is a one to one relation between the optimal policy and the value function, but unfortunately the value function does not possess a closed form representation under our maintaned assumption. Despite this, it is possible to obtain a few insights from the first order condition. Many result rests on the fact that value function is increasing over Γ_N. This is proved later on in Proposition 12. First, the optimal policy is never zero: no matter what the state, the planner can always make its expected cost a little better by supplying some policy.

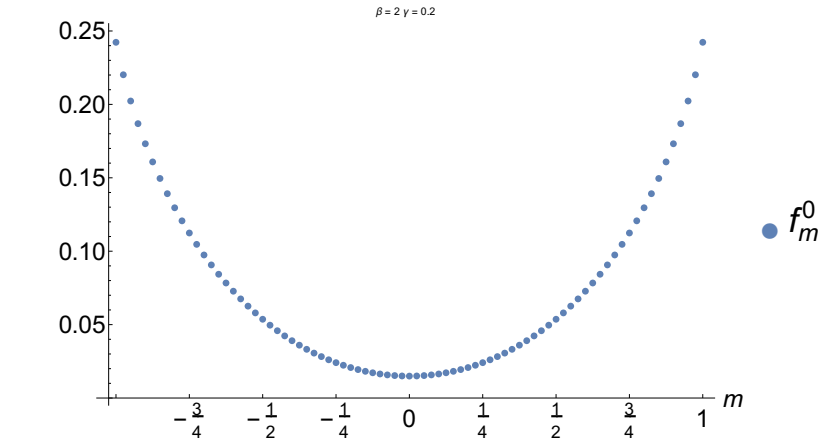
Proposition 10 (Optimal policy is strictly negative) *For $\eta = -1 (\eta = 1)$ the optimal policy $\bar{h}(m)$ is strictly negative (positive) for all m.*

Proof: If V_m is stricly positive and increasing in m, it follows that any h satisfying (5.3) has to be strictly smaller than zero, given that $\partial_h r_m^+ > 0$ and $\partial_h r_m^- < 0$ □ .

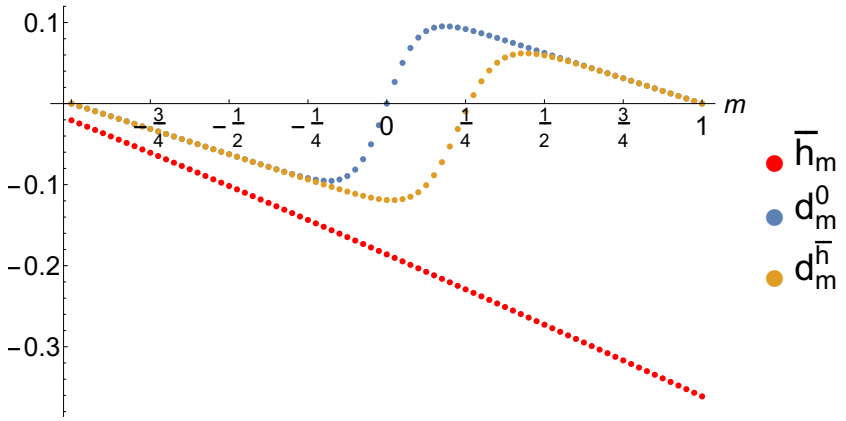
In the following we always assume that the planner target is $\eta = -1$. The optimal policy has several phases which depending on the values of the rationality parameter and the strenght of the coordination motive, which relate to the shape of the stationary distribution when no policy is supplied and in particular whether there are multiple long lived equilibria.

5.2.1 Monotonic case

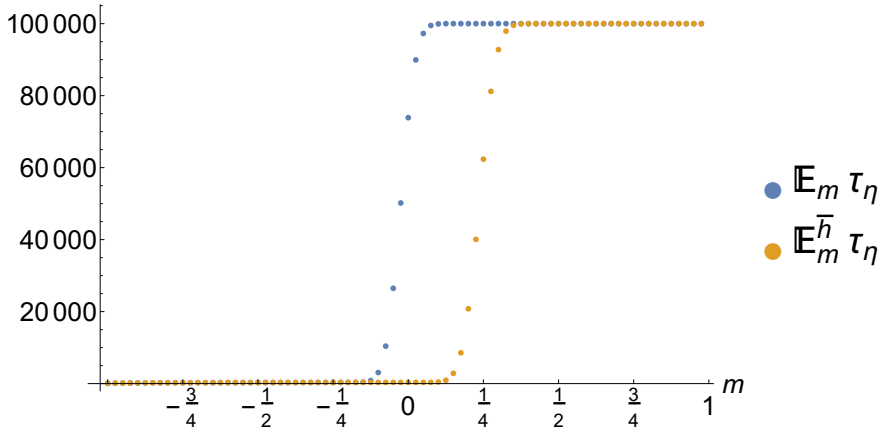
When the coordination motiv γ is low the long period behavior of the chain is driven by the single well potential shown in Fig. 5.2. In this case the policy is monotone and linear and consequently the change in the drift compared with the drift under the null policy, $h_m = 0$ for all m, is just a vertical shift. Since the potential is single welled in this regime the chain spends most of its time near zero and the time to reach the target η is exponential in N. The last plot in Fig. 5.2 shows that this is still the case under the optimal policy, with only little improvement in the expected time required to reach the target when the chain is already close to it.



(a) Stationary distribution under null policy

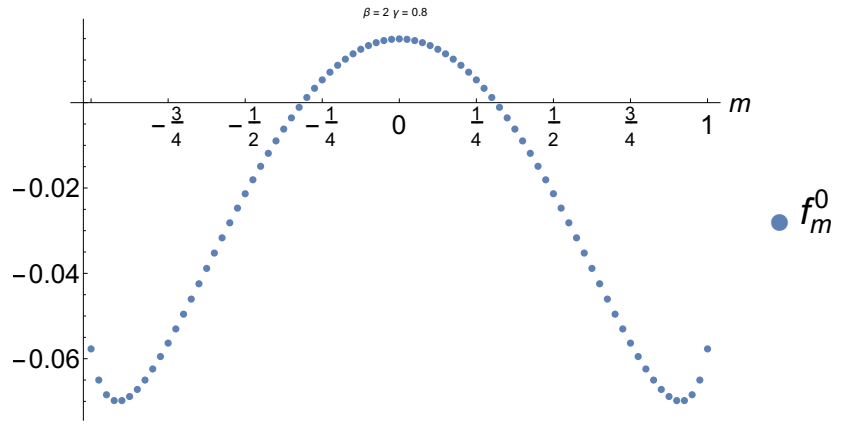


(b) Optimal policy and drift

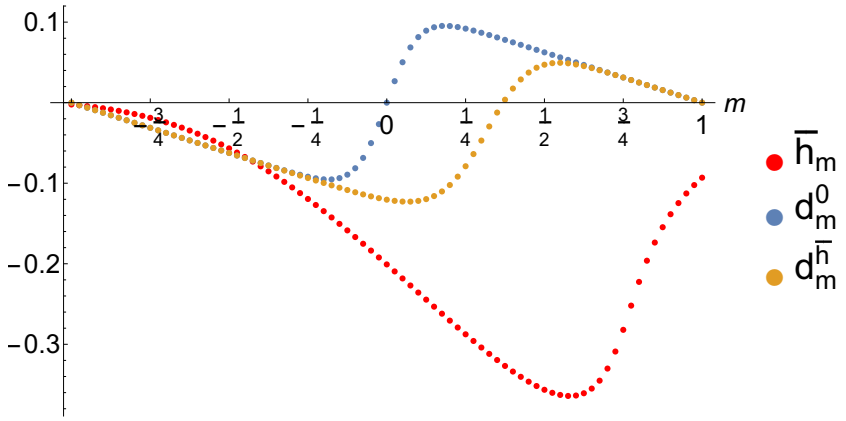


(c) Expected time to reach the planner's target $\eta = -1$

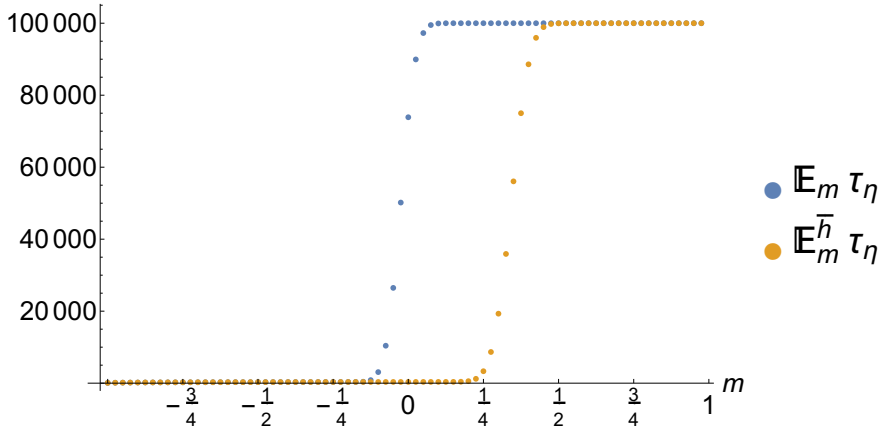
Figure 5.2: Several quantities are plotted here for a population of size $N = 80$. **(a)** The potential function of the lumped stationary distribution μ under the null-policy $h = 0$ with low rationality $\beta = 2$ and weak coordination motiv $\gamma = 0.2$. The distribution is unimodal, meaning that there is a single attracting region for the average population choice m_t. **(b)** In this regime the optimal policy $\bar{h}$ (red) is monotonic, the larger it is in absolute value, the more policy the planner chooses when the process is at position m. The drift under $\bar{h}$ (orange) is only shifted versus the drift under the null-policy (blue). **(c)** The expected hitting time $\mathbb{E}_m[\tau_\eta]$ under the null (blue) and optimal (orange) policy are shown. The effect of the optimal policy is to reduce the expected time to hit the planner target $\eta = -1$ past some critical threshold close to $m = 0$.



(a) Stationary distribution under null policy



(b) Optimal policy and drift



(c) Expected time to reach the target η

Figure 5.3: Several quantities are plotted here for a population of size $N = 80$. **(a)** The potential function of the lumped stationary distribution μ under the null-policy $h = 0$ with low rationality $\beta = 2$ and strong coordination motiv $\gamma = 0.8$. The distribution is bimodal, meaning that there are two regions to which the average population choice m_t is attracted to. **(b)** In this regime the optimal policy $\bar{h}$ (red) is no longer monotonic: gains from of policy increases with the distance from the planner target $\eta = -1$ for a while, but after a certain threshold the pull of the majority on the average choice is so strong, that the return on additional units of policy decreases. The drift experienced by the process under $\bar{h}$ (orange) is shifted versus the drift under the null-policy (blue), so that the region where the process moves on average toward the planner's target is increased. **(c)** The expected hitting time $\mathbb{E}_m[\tau_\eta]$ under the null (blue) and optimal (orange) policies are shown. The effect of the optimal policy is to reduce the expected time to hit the planner target $\eta = -1$ past some critical threshold close to $m = 0$.

5.2.2 Non-monotonic cases

Once the coordination motive parameter is sufficiently large the process has a double well potential and multiple equilibria. The optimal policy is no longer monotone, peaking after $m = 0$, the point where fifty percent of the population has adopted a position opposite to the planner's target. For the parameter of Fig. 5.3, $\bar{h}$ is sufficiently strong to make the drift change sign only once, meaning that the stationary distribution has a single welled potential with a unique equilibrium.

With large coordination and low propensity of mistakes (low β) the non monotonicity becomes abrupt. The behavior of the chain under the null policy, as shown in Fig. 5.4, is that there are two wells whose bottom is located at the edges of the lattice, meaning that the attracting points are the two position -1 and 1. Attraction from these two is so strong that, once the chain is close to $-\eta$, the optimal policy drops suddenly close to zero. The intuition is that the gains from an additional unit of optimal policy are very tiny compared to its cost. This reflects itself in the fact that the value function is no longer convex in the policy parameter, presenting multiple minima that satisfy Eq. (5.3). In the large β regime the global and local minima swap places for some m, see Figure 5.5.

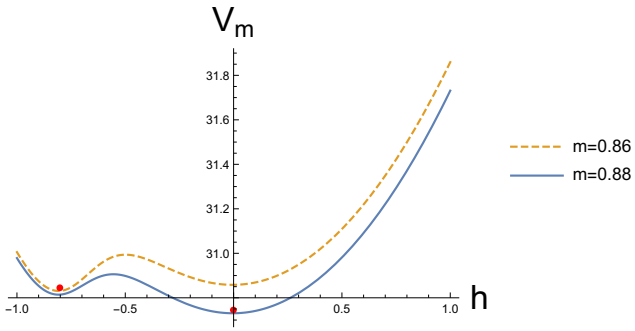
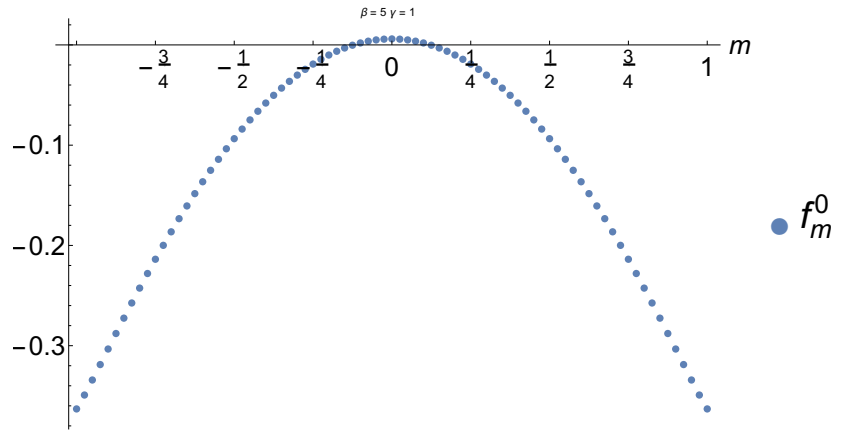


Figure 5.5: Two slices of the value function V_m along the h dimension in the large β and γ regime. The value of h that minimizes the function – marked by a red dot – jumps with m.

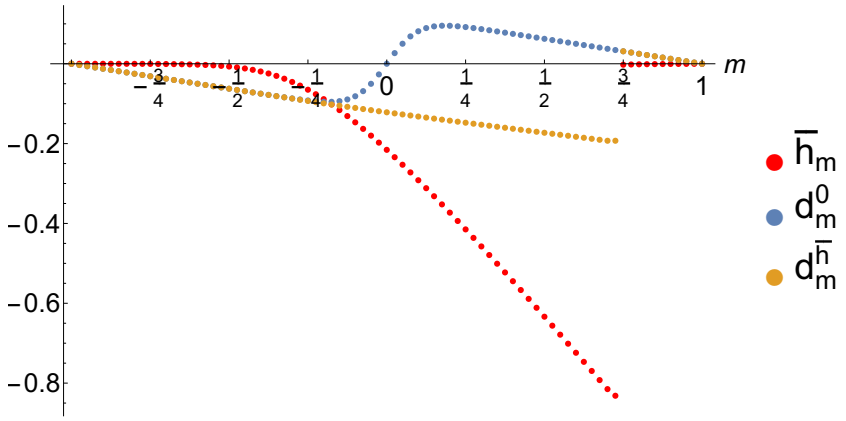
The marginal cost of policy c has an important impact on the value function. Indeed, marginal costs make the difference between the value function presenting a single or a multiplicity of minima along in the choice of policy. We can see this by graphing the following approximation of the second partial derivative of V_m in h:

$$\partial_h^2 V_m = c + \delta(\partial_h^2 r_m^+ \Delta V_{m'} - \partial_h^2 r_m^- \Delta V_m) \approx c + \delta(\partial_h^2 r_m^+ - \partial_h^2 r_m^-). \tag{5.4}$$

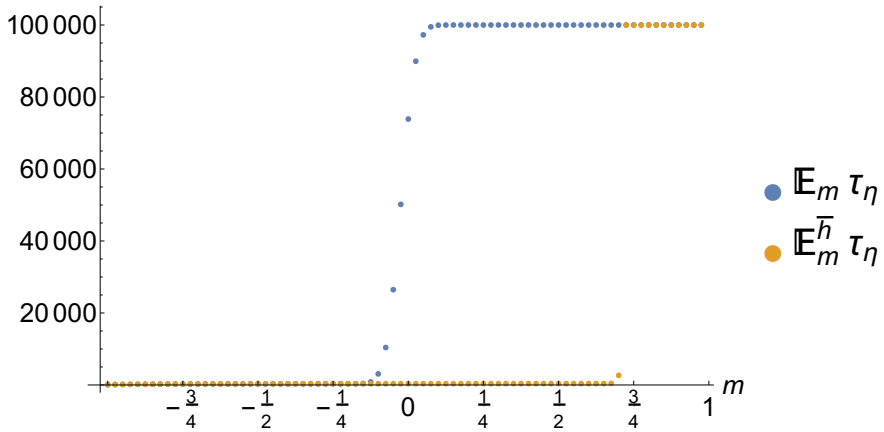
For small values of c there are regions both of convexity and concavity the red



(a) Stationary distribution under null policy



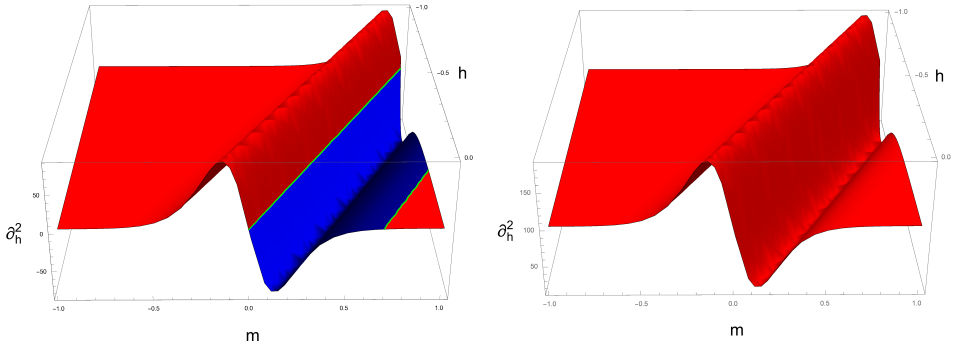
(b) Optimal policy and drift



(c) Expected time to reach the target η

Figure 5.4: Several quantities are plotted here for a population of size $N = 80$. **(a)** The potential function of the lumped stationary distribution μ under the null-policy $h = 0$ with large rationality parameter $\beta = 5$ and strong coordination motiv $\gamma = 1$. The distribution is bimodal, meaning that there are two regions to which the average population choice m_t is attracted to. **(b)** In this regime the optimal policy $\bar{h}$ (red) is no longer monotonic and display a sharp jump past a threshold: pull from the majority on the average choice becomes so strong, that when majority is reached, the gains from policy become extremely small. The drift under $\bar{h}$ (orange) displays a sharp jump as well. **(c)** The expected hitting time $\mathbb{E}_m[\tau_\eta]$ under the null (blue) and optimal (orange) policies are shown. The effect of the optimal policy is to reduce the expected time to hit the planner target $\eta = -1$ past some critical threshold close to $m = 0$. The policy jump translates to the expected times as well.

and blue region in Figure 5.6a respectively, of the value function. When c becomes extremely large instead the function is always concave in h.



(a) $c = 1$ **(b)** $c = 100$

Figure 5.6: Approximated graph of $\partial_h^2 V_m$

These two extremes examples for the marginal cost have a corresponding consequence for the policy: extremely low cost determine a monotonic policy, Figure 5.7, as predicted by Proposition 11 which provides bounds that relate the marginal costs and the agents parameters.

Proposition 11 (Monotonicity of $\bar{h}$) *Consider the upper bound on $\Delta V_m \leq \nu_{\delta,m'}$ given in Lemma 6 for some $m' > 0$. The optimal policy is not monotone for $c > \frac{\delta\beta}{\gamma}\nu_{\delta,m'}$. Conversely, if $c < \frac{2\delta\beta^2}{N}(1-T_+^2)T_+\phi_m^+$ the policy is monotonic.*

Proof: Consider the first order condition in Eq. (5.3), expressing the derivatives of the probability this becomes

$$-ch^* = \frac{\delta\beta}{2}\left[(1-T_+^2)\phi_m^+\Delta V_{m'} + (1-T_-^2)\phi_m^-\Delta V_m\right] \tag{5.5}$$

where $m' = m + 2/N$ and we use the shorthands

$$T_\pm = \tanh\left[\beta\left(h + \gamma m \pm \frac{\gamma}{N}\right)\right] \ , \ \phi^\pm = (1 \mp m).$$

Non-monotonicity: Figure 5.1 gives the graphical rapresentation of 5.5 and shows that there might be multiple interesection. To prove that the optimal policy is not monotonic, first we show that for some strictly positive value of m the peak of the right hand side of (5.5) lies below the line with slope $-c$, implying that there exists a unique intersection to the right of the maximum. By lemma 5 the maximum of the r.h.s. is in the interval $h^* \in \{-\gamma m - \frac{\gamma}{N}, -\gamma m + \frac{\gamma}{N}\}$, hence we need to show that

for for some $m > 0$

$$-ch^* > \frac{\delta\beta}{2}\left[(1-T_+^2)\phi_m^+\Delta V_{m'} + (1-T_-^2)\phi_m^-\Delta V_m\right].$$

Picking $h^* = -\gamma m + \frac{\gamma}{N}$ which makes the inequality as hard as possible to satisfy $T_+^2 \approx O\left(\frac{1}{N}\right)$ and $T_-^2 = 0$

$$\begin{aligned}\frac{\delta\beta}{2}\left[\left(1-O\left(\frac{1}{N}\right)\right)\phi_m^+\Delta V_{m'} + \phi_m^-\Delta V_m\right] &< \frac{\delta\beta}{2}\left[\Delta V_{m'} + \Delta V_m\right] \\ &< \frac{\delta\beta}{N}\nu_{\delta,m'} < c\left(\gamma m - \frac{\gamma}{N}\right)\end{aligned} \tag{5.6}$$

reasoning: for low enough c unique intersection to the left. As c raises to the point where the ineq above hold either the solution remains on the left intersection, in which case it passed the max and went past it, the policy is not monotonic; or the solution jumped to the rightmost intersection, hence the policy is not monotonic. *Monotonicity*: Monotonicity is guaranteed if we ensure that the slope of the l.h.s. is always below the slope of the r.h.s. (i.e. take the second derivative of the foc and discard pieces). Taking the derivative of the r.h.s. in h, we need to ensure that the slope is lower then the slope after h^*, that is that $-c$ is larger

$$-c > -\delta\beta^2\left[(1-T_+^2)T_+\phi_m^+\Delta V_{m'} + (1-T_-^2)T_-\phi_m^-\Delta V_m\right].$$

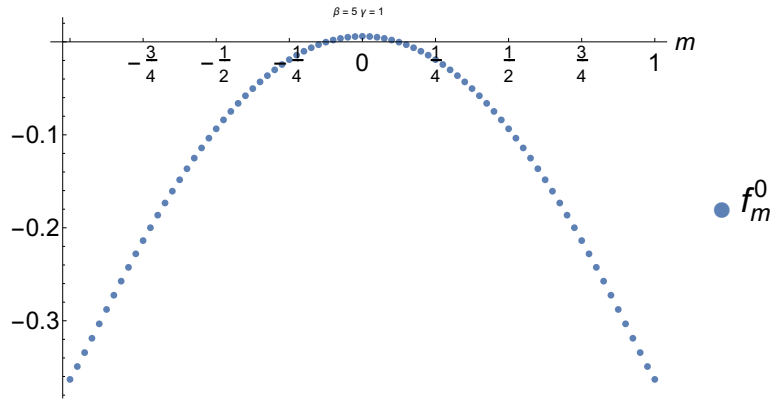
Using the lower bounds on ΔV_m given in Eq. 5.9

$$\begin{aligned}c &< 2\frac{\delta\beta^2}{N}\left[(1-T_+^2)T_+\phi_m^+\right] \\ &< \frac{\delta\beta^2}{N}\left[(1-T_+^2)T_+\phi_m^+ + (1-T_-^2)T_-\phi_m^-\right] \\ &< \delta\beta^2\left[(1-T_+^2)T_+\phi_m^+\Delta V_{m'} + (1-T_-^2)T_-\phi_m^-\Delta V_m\right]\end{aligned}$$

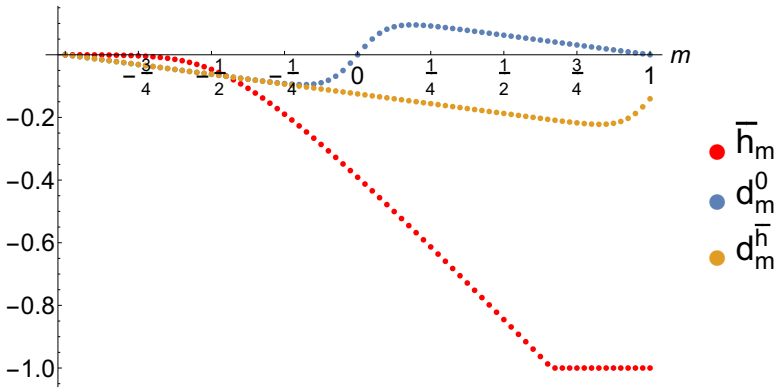
□.

5.2.3 Optimal policy location

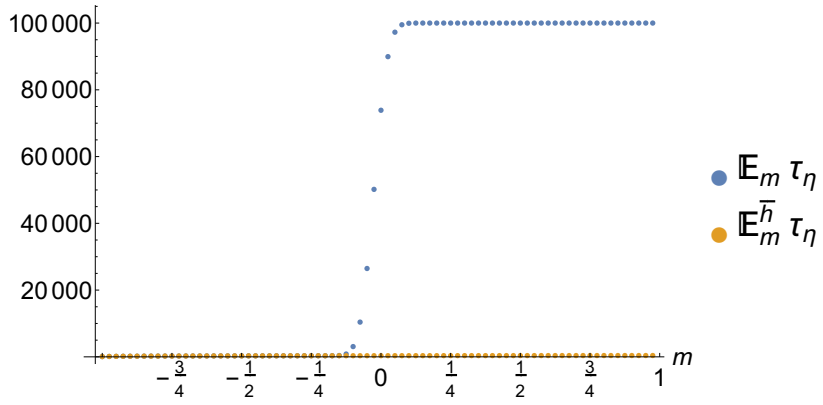
From the various plot of the optimal policy it emerges that, when the policy is not monotone, there exists a critical point after which the attraction from the position adversed by the planner becomes so strong that that it's not convenient anymore to invest in policy, that is, returns from additional units of policy start to decline. In this section we show that this critical point always lies to the right of zero. This means, that when the planner is targeting $\eta = -1$, happens before half of the



(a) Stationary distribution under null policy

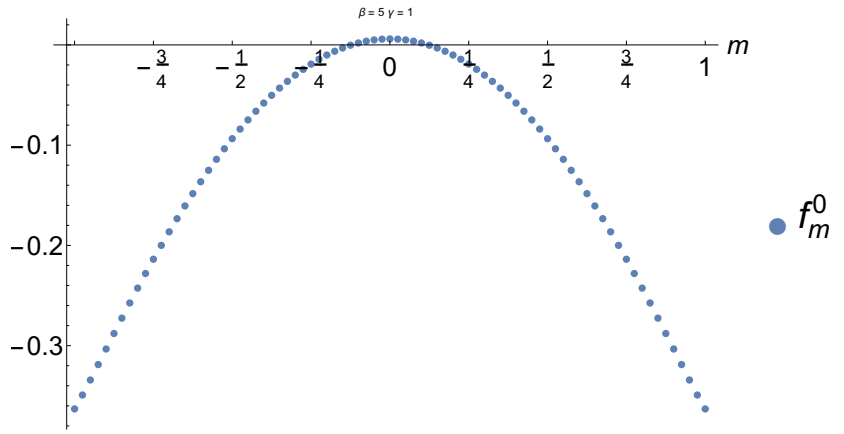


(b) Optimal policy and drift

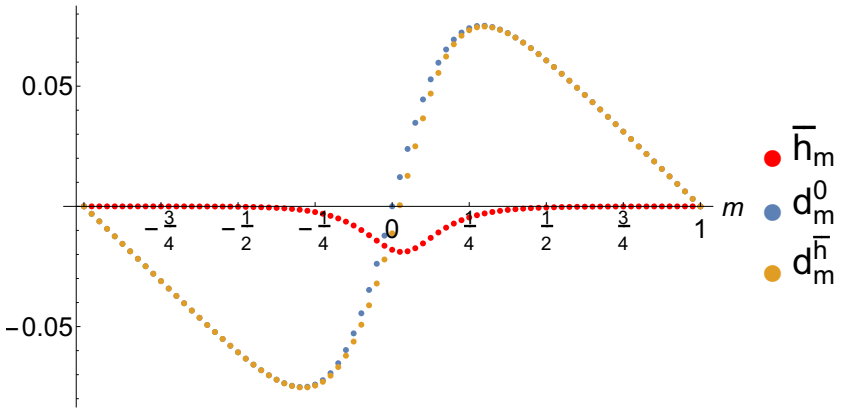


(c) Expected time to reach the target η

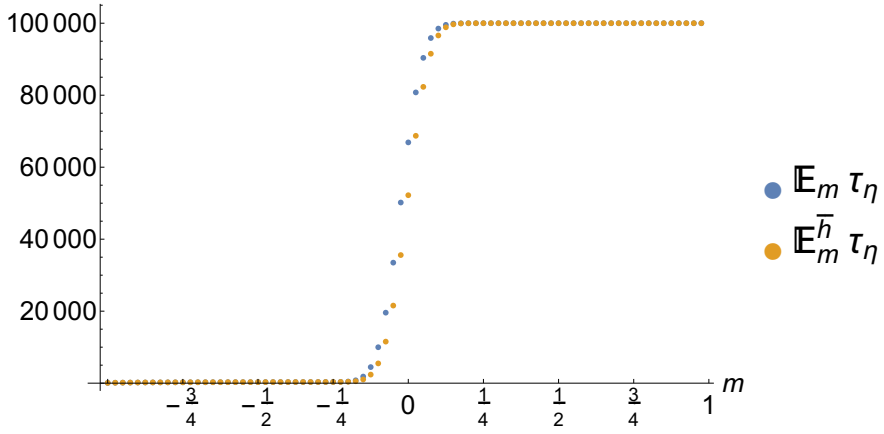
Figure 5.7: Several quantities are plotted here for a population of size $N = 80$ with low marginal costs of policy $c = 0.1$. **(a)** The potential function of the lumped stationary distribution μ under the null-policy $h = 0$ with large rationality parameter $\beta = 5$ and strong coordination motiv $\gamma = 1$. The distribution is bimodal, meaning that there are two regions to which the average population choice is attracted to. **(b)** In this regime the optimal policy $\bar{h}$ (red) is still monotonic and at times maxed to -1. Low marginal costs of policy imply that returns on policy are always large. The drift under $\bar{h}$ (orange) is always negative, hence the process on average always drifts toward the planner's target. Drift being zero under $\bar{h}$ only at -1 suggests the second attracting region disappears under the optimal policy. **(c)** The expected hitting time $\mathbb{E}_m[\tau_\eta]$ under the null (blue) and optimal (orange) policies are shown. The effect of the optimal policy is to reduce the expected time to hit the planner target $\eta = -1$ over the whole state space.



(a) Stationary distribution under null policy



(b) Optimal policy and drift



(c) Expected time to reach the target η

Figure 5.8: Several quantities are plotted here for a population of size $N = 80$ with high marginal costs of policy $c = 100$. **(a)** The potential function of the lumped stationary distribution μ under the null-policy $h = 0$ with large rationality parameter $\beta = 5$ and strong coordination motiv $\gamma = 1$. The distribution is bimodal, meaning that there are two regions to which the average population choice is attracted to. **(b)** In this regime the optimal policy $\bar{h}$ (red) is significant only in the most critical region right after a majority for the planner choice is lost. After this, the pull from the opposing majority quickly becomes too strong and the policy is quickly tapered down. The drift under $\bar{h}$ (orange) is just slightly shifted compared with the null policy. **(c)** The expected hitting time $\mathbb{E}_m[\tau_\eta]$ under the null (blue) and optimal (orange) policies are shown and are not majorily changed by the optimal policy.

population has changed favorably its position. Hence, the planner would want to intervene, and it's actually better off, when a small majority of the population does not share the planner position.

First we need to show that the value function is increasing over the state space of the chain. This is a desirable (and intuitively reasonable) property of the value function but is usually proved either relying on stochastic dominance or on monotonicity of the optimal policy. Neither of these assumption is true in this context and coupling techniques are needed.

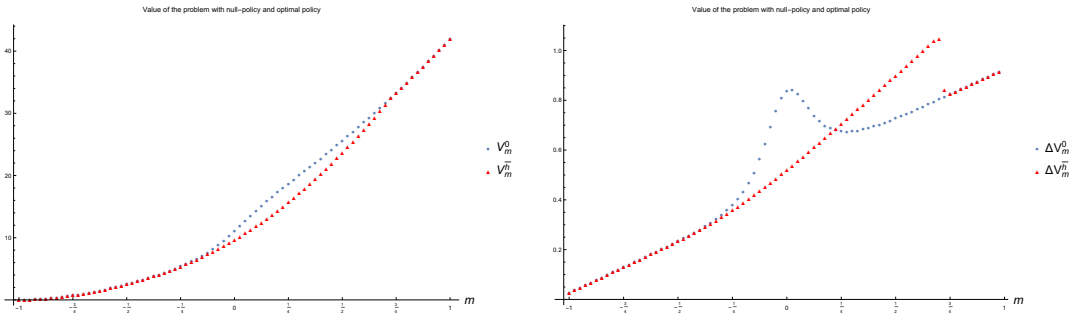


Figure 5.9: Numerical plot of V_m and ΔV_m under the null and optimal policy with parameters $N = 80, \beta = 5, \gamma = 1, \delta = 0.9$ and $c = 1$.

The challenge of the proof lies in the need to compute expectations of the chain under the optimal policy $\bar{h}$. By constructing a appropriate joint distribution of two copies of the chain, i.e. a coupling, we can derive monotonic relationship between the lumped measures which then extend to the expectations by Strassen's theorem (see thm.6).

Proposition 12 (V_m is increasing) *$V_m > V_n$ for all $m > n \in \Gamma_N$.*

Proof:

Coupling of the original process

Consider two instances of the non lumped chain $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$, defined in section 3.1, with initial conditions $\boldsymbol{x} \geq \boldsymbol{y}$ where the inequality is pointwise. Both chain live in the same probability space $(\Lambda_N, \sigma(\Lambda_N), \mathbb{P})$.

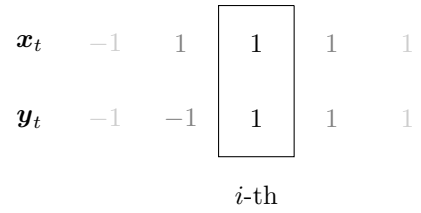


Figure 5.10

A coupling is a joint distribution between the two chain such that the marginal distribution with respect to each one is the marginal distribution of the two original chains. Define τ to be the first time the two chains are equal

$$\tau = \inf_t \{t \geq 1 : \boldsymbol{x}_t = \boldsymbol{y}_t\},$$

and denote $\boldsymbol{x}^i$ the configuration which is identical to $\boldsymbol{x}$ but has the sign of the i-th element changed. The probability of agent i revising[1] his current choice is

$$P_{\boldsymbol{x}\boldsymbol{x}^i} = \left[1 + e^{-2\beta(-x_i h - x_i m_{\boldsymbol{x}} + \frac{\gamma}{N})}\right]^{-1} = \frac{1}{2}\left\{1 - \tanh\left[\beta\left(x_i h + x_i \gamma m_{\boldsymbol{x}} - \frac{\gamma}{N}\right)\right]\right\}.$$

The coupling is defined as follow: start two copies of the chain $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ at the intial condition $\boldsymbol{x} \geq \boldsymbol{y}$. If $t \geq \tau$ they move together according to the transition probability matrix P; for any $t < \tau$ select a random coordinate i, then sort a uniform random variable $Z \sim U[0,1]$. Conditional on the initial state of the joint chain (x_i, y_i) and the realization of Z update the state of both chains according to the diagram in Fig. 5.11.

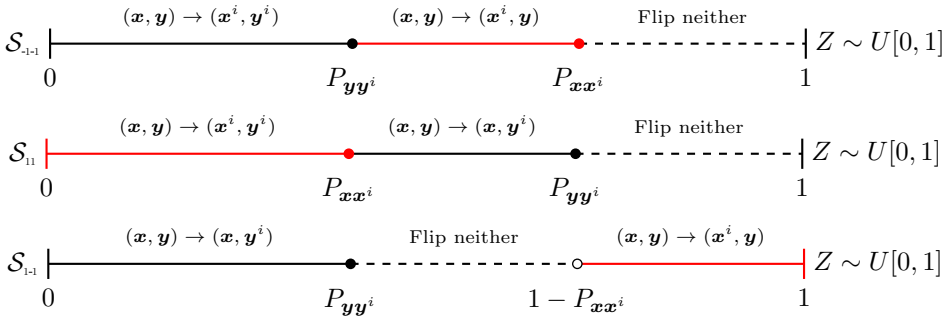


Figure 5.11: Schematic of the coupling, according to the region where the uniform random draw Z falls the $i-th$ component of $\boldsymbol{x}_t$ or $\boldsymbol{y}_t$ changes sign. Note that until the two chains are identical the coupling ensures that $\boldsymbol{x}_t \geq \boldsymbol{y}_t$.

The joint distribution of the coupling $\mathcal{C}$, assigning probability of moving from

[1] Using $m_{\boldsymbol{x}} = \sum_i x_i$.

a pair of configurations $(\boldsymbol{x}, \boldsymbol{y})$ to some other pair $(\boldsymbol{v}, \boldsymbol{z})$ is given by:

$$\mathcal{C}_{\boldsymbol{xy}}(\boldsymbol{v}, \boldsymbol{z}) = \frac{1}{N} \begin{cases} 0 & \text{if } ||\boldsymbol{x} - \boldsymbol{v}|| > 2 \text{ or } ||\boldsymbol{y} - \boldsymbol{z}|| > 2 \\ \sum_i \mathcal{S}_{\boldsymbol{xy}}(\boldsymbol{v}, \boldsymbol{z}) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{v}, \boldsymbol{z}) \\ \mathcal{S}_{\boldsymbol{xy}}(\boldsymbol{v}, \boldsymbol{z}) & \text{othw.} \end{cases}$$

where the terms $\mathcal{S}_{\boldsymbol{xy}}(\boldsymbol{v}, \boldsymbol{z})$ are reported in Table 5.1. This construction means that the joint distribution ensures that the monotonicity of the chains is preserved until the chains meet at some random time τ.

$\mathcal{S}_{11}(\cdot)$

	$\boldsymbol{y}$	$\boldsymbol{y}^i$
$\boldsymbol{x}$	$1 - P_{\boldsymbol{yy}^i}$	$P_{\boldsymbol{yy}^i} - P_{\boldsymbol{xx}^i}$
$\boldsymbol{x}^i$	0	$P_{\boldsymbol{xx}^i}$

$\mathcal{S}_{\text{-}1\text{-}1}(\cdot)$

	$\boldsymbol{y}$	$\boldsymbol{y}^i$
$\boldsymbol{x}$	$1 - P_{\boldsymbol{xx}^i}$	0
$\boldsymbol{x}^i$	$P_{\boldsymbol{xx}^i} - P_{\boldsymbol{yy}^i}$	$P_{\boldsymbol{yy}^i}$

$\mathcal{S}_{1\text{-}1}(\cdot)$

	$\boldsymbol{y}$	$\boldsymbol{y}^i$
$\boldsymbol{x}$	$1 - P_{\boldsymbol{yy}^i} - P_{\boldsymbol{xx}^i}$	$P_{\boldsymbol{yy}^i}$
$\boldsymbol{x}^i$	$P_{\boldsymbol{xx}^i}$	0

Table 5.1: Joint distrubtion under the coupling conditional on having picked the i-th

Now we show that the joint distribution is well defined. Starting from a pair of configurations such that $(x_i, y_i) = (1, 1)$ then

$$\begin{aligned} 0 \leq P_{\boldsymbol{yy}^i} - P_{\boldsymbol{xx}^i} &= \tanh\left[\beta\left(-y_i h - y_i \gamma m_{\boldsymbol{y}} + \frac{\gamma}{N}\right)\right] - \tanh\left[\beta\left(-x_i h - x_i \gamma m_{\boldsymbol{x}} + \frac{\gamma}{N}\right)\right] \\ &= \tanh\left[\beta\left(-h - \gamma m_{\boldsymbol{y}} + \frac{\gamma}{N}\right)\right] - \tanh\left[\beta\left(-h - \gamma m_{\boldsymbol{x}} + \frac{\gamma}{N}\right)\right] \end{aligned}$$

which follows from $m_{\boldsymbol{x}} \geq m_{\boldsymbol{y}}$ and similarily for $(x_i, y_i) = (-1, -1)$. When configurations are such that $(x_i, y_i) = (1, -1)$

$$\begin{aligned} 0 &\leq 1 - P_{\boldsymbol{yy}^i} - P_{\boldsymbol{xx}^i} \\ 0 &\geq \tanh\left[\beta\left(h + \gamma m_{\boldsymbol{y}} + \frac{\gamma}{N}\right)\right] + \tanh\left[\beta\left(-h - \gamma m_{\boldsymbol{x}} + \frac{\gamma}{N}\right)\right] \\ \tanh\left[\beta\left(-h - \gamma m_{\boldsymbol{x}} + \frac{\gamma}{N}\right)\right] &\leq \tanh\left[\beta\left(-h - \gamma m_{\boldsymbol{y}} - \frac{\gamma}{N}\right)\right] \\ -h - \gamma m_{\boldsymbol{x}} + \frac{\gamma}{N} &\leq -h - \gamma m_{\boldsymbol{y}} - \frac{\gamma}{N} \\ \frac{2}{N} &\leq m_{\boldsymbol{x}} - m_{\boldsymbol{y}} \end{aligned}$$

which always holds since $\boldsymbol{x} \geq \boldsymbol{y}$.

For this joint distribution to be a coupling the marginal distribution of the chain must return the distribution of the original chains $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$. This is immediatly obvious by marginalization of Table 5.1. Further, the coupling is successful, that is, the probability that the chains meet eventually is one. This is trivial thanks to the fact that both chains are irreducibile.

Having shown that a successful coupling exists, by Strassen's theorem (see Theorem 6) this means that their probability measures conditioned on the starting configuration are increasing in the initial configuration $\mathbb{P}_{\boldsymbol{x}} \geq \mathbb{P}_{\boldsymbol{y}}$ and so are their conditional expectations.

Coupling of the lumped process

Under the monotonic map $m(\boldsymbol{x}) = \sum_i x_i/N$ the chain $m_t = m(\boldsymbol{x}_t)$ and $n_t = m(\boldsymbol{y}_t)$ are also coupled, their joint distribution is given in Table 5.2

$m_{\boldsymbol{x}} \setminus m_{\boldsymbol{y}}$	$+$	$-$	$\cdot$
$+$	$\varphi^- P_{\boldsymbol{y}\boldsymbol{y}^i}$	0	$\varphi^-(P_{\boldsymbol{x}\boldsymbol{x}^i} - P_{\boldsymbol{y}\boldsymbol{y}^i})$
$-$	0	$\varphi^+ P_{\boldsymbol{x}\boldsymbol{x}^i}$	$\varphi^\pm P_{\boldsymbol{x}\boldsymbol{x}^i}$
$\cdot$	$\varphi^\pm P_{\boldsymbol{y}\boldsymbol{y}^i}$	$\varphi^+(P_{\boldsymbol{y}\boldsymbol{y}^i} - P_{\boldsymbol{x}\boldsymbol{x}^i})$	$\varphi^+(1 - P_{\boldsymbol{y}\boldsymbol{y}^i}) + \varphi^-(1 - P_{\boldsymbol{x}\boldsymbol{x}^i}) + \varphi^\pm(1 - P_{\boldsymbol{y}\boldsymbol{y}^i} - P_{\boldsymbol{x}\boldsymbol{x}^i})$

Table 5.2: Joint distrubtion under the coupling conditional on having picked the i-th. Where φ denotes the number of sites of a configuration with a certain starting condition : $\varphi^- = |\{i : (x_i, y_i) = (-1, -1)\}|$, $\varphi^+ = |\{i : (x_i, y_i) = (1, 1)\}|$, $\varphi^\pm = |\{i : (x_i, y_i) = (1, -1)\}|$
.

Consider two chain m_t, n_t with the initial conditions $m > n$ which are generated by the underlying coupled chains $\boldsymbol{x}_t, \boldsymbol{y}_t$. Recall the definition of the value function (5.1) as expected discounted costs

$$V_m := \inf_h \mathbb{E}^h \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \,\middle|\, m_0 = m \right] = \\ = \inf_h \mathbb{E}^h_m \left[\sum_{t=0}^{\infty} \delta^t C(m_t, h_t) \right],$$

we will make use of the coupling expectations, which are monotone by Strassen's theorem, to show that if $m > n$, then $V_m > V_n$.

So far, in order for the coupling to work, we assumed that at each time the policy parameter was constant $h_t = h$. This can be relaxed as long as both chains have the same h when updated. We are going to pick $h_t = \bar{h}(m_{t-1})$ for both chains, that is we couple both chain using the probability that depends on the optimal

policy for the chain started with a higher initial condition. The sequence h_t is well defined by to the initial condition and by the fact that the optimal policy is also stationary[2]. Denote the coupling expectation $\mathbb{E}_{\mathcal{C}}^{\boldsymbol{h}}$:

$$\begin{aligned} V_m &= \inf_{\boldsymbol{h}} \mathbb{E}_{\mathcal{C}}^{\boldsymbol{h}} \sum_{t\geq 0} \delta^t \Big[h_t^2 + (m_t - \eta)^2 \Big] \\ &= \inf_{\boldsymbol{h}} \mathbb{E}_{m}^{\boldsymbol{h}} \sum_{t\geq 0} \delta^t \Big[h_t^2 + (m_t - \eta)^2 \Big] \\ &= \quad \mathbb{E}_{m}^{\bar{\boldsymbol{h}}} \sum_{t\geq 0} \delta^t \Big[\bar{h}^2(m_t) + (m_t - \eta)^2 \Big]. \end{aligned} \tag{5.7}$$

Here $\boldsymbol{h}$ is just compact notation to denote any sequence of controls $\{h_t\}_{t\geq 0}$. The infimum is taken over all possible sequences and under the coupling expectation is the same as the one under the marginal expectation. Now consider the value $\widetilde{V}_n$ of the chain starting at n and using the sequence $h_t = \bar{h}(m_{t-1})$. Since this sequence is optimal for the chain started in m it is suboptimal when the chain starts in n instead, therefore:

$$\begin{aligned} \widetilde{V}_n &= \mathbb{E}_{\mathcal{S}}^{\bar{h}} \sum_{t\geq 0} \delta^t \Big[\bar{h}(m_t)^2 + (n_t - \eta)^2 \Big] \\ &= \sum_{t\geq 0} \delta^t \Big\{ \mathbb{E}_{m}^{\bar{h}}[\bar{h}(m_t)^2] + \mathbb{E}_{n}^{\bar{h}}[(n_t - \eta)^2] \Big\} \\ &\geq V_n \end{aligned} \tag{5.8}$$

bounding V_n from below. The second line follows from marginalization and the last line follows because V_n is an infimum over all possible control sequences.

$$\begin{aligned} V_m - V_n &\geq V_m - \widetilde{V}_n \\ &= \sum_{t\geq 0} \delta^t \Big\{ \mathbb{E}_{m}^{\bar{h}}[((m_t - \eta)^2] - \mathbb{E}_{n}^{\bar{h}}[(n_t - \eta)^2] \Big\} \\ &= (m-\eta)^2 - (n-\eta)^2 + \sum_{t>0} \delta^t \Big\{ \mathbb{E}_{m}^{\bar{h}}[(m_t - \eta)^2] - \mathbb{E}_{n}^{\bar{h}}[(n_t - \eta)^2] \Big\} \\ &> (m-\eta)^2 - (n-\eta)^2 > 0 \end{aligned} \tag{5.9}$$

The first inequality follows from (5.8). The first equality follows from the fact that the coupling expectation over $\bar{h}(m_t)$ reduces to the marginal expectation of m_t and the policy terms cancel out, since they are the same. The last step follows from the fact that $\mathbb{E}_m \geq \mathbb{E}_z$ by Strassen's theorem and the existence of a succesful coupling

[2] That is, it is constant in time and only depends on the current state of the chain

and the monotonicity of $(m-\eta)^2$ for $\eta = -1$. □

We are now ready to prove that the maximum amount of policy is applied before half of the population shares the preferred position of the planner.

Proposition 13 (Maximum pressure is applied past $m = 0$) *Consider the infinite horizon planner problem, with parameters $\beta > 1$ and $0 < \gamma < 1$. If $\delta > \frac{\gamma}{\beta(2-O(1/N))}$, then $\arg\min_m \bar{h}_m > 0$.*

Proof: Consider the implicit function describing the first order condition of the planner problem in a neighborhood of the optimal value $\bar{h}$:

$$F(h,m) = ch + \delta\left[\partial_h r_m^+ \Delta V_{m+1} - \partial_h r_m^- \Delta V_m\right] = 0.$$

For a sufficiently large population N we can consider the derivative in m, then by the implicit function theorem, the change in the optimal policy when m changes are given by

$$\frac{\partial \bar{h}}{\partial m}\Big|_{m=0} = -\frac{\delta\beta}{2}\frac{\left(1-T_-^2\right)\left(1+2\beta\gamma T_-\right)\Delta V_m - \left(1-T_+^2\right)\left(1+2\beta\gamma T_+\right)\Delta V_{m+1} + O\left(\frac{1}{N}\right)}{2c + \delta\beta^2\left((1-T_-^2)T_-\Delta V_m - (1-T_+^2)T_+\Delta V_{m+1}\right)} \tag{5.10}$$

Recall, that for target $\eta = -1$ the planner always chooses a negative policy (see Proposition 10). We would like to show that at $m = 0$ the derivative is negative, hence the policy has to increase for higher m in absolute value, implying that the maximum is located to its right.

Denominator

Since $\bar{h} < 0$ then $0 < T_- < 1$. It follows that the denominator is always positive if $T_+ < 0$ which is the case under the condition of lemma 4, since $-\bar{h} > \frac{\gamma}{N}$. Even when conditions do not hold, bounds on $\Delta V. \approx O(1/N)$, so for any c there's an N large enough so that the denominator is strictly positive. In particular,

$$N > \sqrt{\frac{\delta\beta^2}{2c}K}$$

where K is some positive constant.

Numerator

The $O(1/N)$ term is

$$\partial_h r_m^+(\Delta V_{m''} - \Delta V_{m'}) - \partial_h r_m^-(\Delta V_{m'} - \Delta V_m)$$

the order following from upper and lower bounds on ΔV_m.
The numerator is positive if

$$\frac{(1-T_+^2)\,(1+2\beta\gamma T_+)}{(1-T_-^2)\,(1+2\beta\gamma T_-)} < \frac{\Delta V_m}{\Delta V_{m+1}} \tag{5.11}$$

Indeed, for a sufficiently large β the following holds

$$\frac{(1-T_+^2)}{(1-T_-^2)}\frac{(1+2\beta\gamma T_+)}{(1+2\beta\gamma T_-)} \;< (1+O(\beta\gamma))\,O\left(\frac{1}{2\beta\gamma}\right) < \frac{\Delta V_m}{\Delta V_{m+1}} \tag{5.12}$$

Given $T_+ < 0$ since $\bar{h} < -\frac{\gamma}{N}$, then

$$\begin{aligned}\frac{1+2\beta\gamma T_+}{1+2\beta\gamma T_-} \;&< \frac{1}{1+2\beta\gamma T_-} \\ &< \frac{1}{2\beta\gamma T_-} \;\approx O\left(\frac{1}{2\beta\gamma}\right)\end{aligned} \tag{5.13}$$

Using the bounds on ΔV_m the right side is a constant which is independent of N, meaning that for a large enough β and consequently N the IFT derivative is strictly positive, so that the absolute value of the policy at zero is increasing in m. It can be that, for a particular N there is no zero, in which case is enough to add one to N. Note that m is, for a finite N, discrete and so is the optimal policy. Since we are assuming a large N, we also assume the implicit function theorem derivative is a good approximation of the variation in the optimal policy when parameters changes □.

5.2.4 Stationary distribution under the optimal policy

Lastly, we provide a calculation based on reversibility for the stationary distribution under a state-dependent policy. This result is unfortunately not as elegant as the stationary distribution presented in Eq. (2).

Proposition 14 (Stationary distribution under optimal policy) *The stationary distribution under the optimal policy is given by*

$$\bar{\mu}[m_i] = \frac{e^{f_i}}{\sum_{j=0}^{N} e^{f_i}}$$

with

$$f_i = g_i + \sum_{\ell=1}^{i} 2\beta(\bar{h}_{m_\ell} + \gamma m_\ell) + I_{\ell\ell} + C_{\ell\ell}$$

for $i \neq 0$ *and* $f_0 = 0$. *Here:*

$$g_i = \beta(\bar{h}_{m_i} + \gamma m_i + \bar{h}_{m_0} + \gamma m_0) + I_{i1} + C_{i1}$$

$$C_{jk} = \ln\cosh[\beta(\bar{h}_{m_j} + \gamma m_j - \gamma/N)] - \ln\cosh[\beta(\bar{h}_{m_k} + \gamma m_k + \gamma/N)$$

$$I_{jk} = \ln(1 - m_j) - \ln(1 + m_k)$$

Proof:

Under any stationary policy the chain posess a unique stationary distribution $\bar{\mu}[m_i]$ and it is reversible respect to it, so for any m_i

$$\bar{\mu}[m_i] = \frac{R_{m_{i-1},m_i}}{R_{m_i,m_{i-1}}}\bar{\mu}[m_{i-1}] = \frac{r^+_{m_{i-1}}}{r^-_{m_i}}\bar{\mu}[m_{i-1}]$$

substituting recursively we get

$$\bar{\mu}[m_i] = \prod_{\ell=1}^{i} \frac{r^+_{m_{\ell-1}}}{r^-_{m_\ell}}\bar{\mu}[m_0]$$

and using the normalization $\sum_i \bar{\mu}[m_i] = 1$ we obtain

$$\bar{\mu}[m_0] = \sum_{j=0}^{N}\prod_{\ell=1}^{i} \frac{r^+_{m_{\ell-1}}}{r^-_{m_\ell}}$$

The transition probabilities can be written as

$$r^\pm_m = \frac{(1 \mp m)}{2}\frac{e^{\pm\beta(\bar{h}_m + \gamma m \pm \gamma/N)}}{\cosh(\beta(\bar{h}_m + \gamma m \pm \gamma/N))}$$

which yield the specific form of f_ℓ □.

5.3 Lemmas

This section gives a few lemma that have been omitted from the main discussion for readability. The most notable ones are Lemma 6 and Lemma 7 which provide upper bounds on the variation of the value function ΔV_m and lower bounds on the optimal policy $\bar{h}$. I have derived this lemmas myself, though I imagine these to be either known or immediate consequences of theorem presented here.

Lemma 3 (Myopic planner value function)

$$V_m > V_m^0$$

and

$$\Delta V_m > \Delta V_m^0$$

Proof: The first claim follows from the definition of the terminal payoff $V_m^0 = (m-\eta)^2$ and the definition of the payoff of the infinite horizon planner by inspection.

For the second claim recall the coupling inequality which bounds ΔV_m from below:

$$\begin{aligned} \Delta V_m &> V_m - \widetilde{V}_n \\ &= (m-\eta)^2 - (n-\eta)^2 + \mathbb{E}_{\mathcal{S}}[\ldots] \\ &> \Delta V_m^0 = (m-\eta)^2 - (n-\eta)^2 \\ &= (m-n)(m+n-2\eta) > \frac{2}{N} \end{aligned} \tag{5.14}$$

The strictness of both inequality follows from the same consideration. The value of the chains at t=0 at different initial conditions under the coupling differ by at least $\Delta(m-\eta)$. The expectation term is strictly positive by the monotonicity of the measures and by the fact that the probability of the chains couple successfuly is positive for $t > 1$, formally $P(m_t = z_t) > 0$ for $t > 1$ $\square$.

Lemma 4 (Myopic planner policy is always smaller than inifinite horizon planner)

$$|\bar{h}_m^1| < |\bar{h}_m^\infty|$$

Proof:
We want to show that

$$-\bar{h}_m^1 = \arg\min[h^2 + \delta \sum_k P_{mk} V_k^0] < \arg\min[h^2 + \delta \sum_k P_{mk} V_k] = -\bar{h}_m^\infty$$

From the first order condition (5.3) we get

$$\partial_h r_m^+ \Delta V_l^0 - \partial_h r_m^- \Delta V_m^0 < \partial_h r_m^+ \Delta V_l - \partial_h r_m^- \Delta V_m \tag{5.15}$$

which rewrites to

$$\frac{\partial_h r_m^+}{\partial_h r_m^-} < \frac{\Delta V_l - \Delta V_l^0}{\Delta V_m - \Delta V_m^0} \tag{5.16}$$

the left hand side is negative by inspection and the right hand side is strictly positive by lemma 3.

Lemma 5 (Argmax for conical combination of unimodal functions) *Consider a unimodal function $f(x) : \mathbb{R} \to \mathbb{R}$ with $\hat{x} = \arg\max f(x)$. Now consider the conical combination of shifted functions with $A, B, \epsilon > 0$*

$$F_\epsilon(x) = Af(x + \epsilon) + Bf(x - \epsilon).$$

Then

$$\arg\max F_\epsilon(x) \in \hat{X} = [\hat{x} - \epsilon, \hat{x} + \epsilon].$$

Proof: Proof is by contradiction. Assume the above is false, then there is an

$$\tilde{x} \notin \hat{X} \ : \ F_\epsilon(\tilde{x}) - F_\epsilon(\hat{X}) > 0.$$

Consider $\tilde{x} = \hat{x} + \epsilon + \delta$ for $\delta > 0$ then

$$F_\epsilon(\tilde{x}) - F_\epsilon(x) = A[f(\hat{x} + 2\epsilon + \delta) - f(x + \epsilon)] + B[f(\hat{x} + \delta) - f(x - \epsilon)]$$

picking $x = \hat{x} + \epsilon$ by unimodality and $\hat{x}$ being the maximizer of f we get

$$F_\epsilon(\tilde{x}) - F_\epsilon(x) = A[f(\hat{x} + 2\epsilon + \delta) - f(\hat{x} + 2\epsilon)] + B[f(\hat{x} + \delta) - f(\hat{x})] < 0$$

a contradiction. Similarly for the case $\tilde{x} < \hat{x} - \epsilon$ □.

Lemma 6 (Upper bound for ΔV_m)

$$\Delta V_m < \frac{2}{N} \nu_{\delta,m}$$

Proof:

For $m > n$ using the reversed coupling inequality

$$
\begin{aligned}
\Delta V_m &\leq \widetilde{V}_m - V_n \\
&= \mathbb{E}_{\mathcal{S}}^{\bar{\bar{h}}} \sum_{t\geq 0} \delta^t \left[\bar{h}(n_t)^2 + (m_t - \eta)^2 \right] - \mathbb{E}_{\mathcal{S}}^{\bar{\bar{h}}} \sum_{t\geq 0} \delta^t \left[\bar{h}(n_t)^2 + (n_t - \eta)^2 \right] \\
&= \sum_{t\geq 0} \delta^t \mathbb{E}_{\mathcal{S}}^{\bar{\bar{h}}} \Big[(m_t - \eta)^2 - (n_t - \eta)^2 \Big] \\
&= \sum_{t\geq 0} \delta^t \mathbb{E}_{\mathcal{S}}^{\bar{\bar{h}}} \Big[(m_t - n_t)(m_t + n_t - 2\eta) \Big] \qquad (5.17) \\
&< \sum_{t\geq 0} \delta^t \left[\left(m - n + \frac{4}{N} t \right) 4 \right] = 4 \sum_{t\geq 0} \delta^t \left[\left(\frac{2}{N} + \frac{4}{N} t \right) \right] \\
&= \frac{8}{N} \sum_{t\geq 0} \delta^t \left[(1 + 2t) \right] = \frac{8}{N} \left[\frac{m + n + 2}{1 - \delta} + \frac{2\delta}{1 - \delta^2} \right]
\end{aligned}
$$

Lemma 7 (Lower bound on $\bar{h}$) *For $\delta > \frac{\gamma}{\beta(2+O(1/N))}$ then $\bar{h} < -\frac{\gamma}{N}$*

Proof: First, write out explicitely the first order condition for the optimal policy

$$
-h = \frac{\delta\beta}{4} \left[\phi_m^+ (1 - T_+^2) \Delta V_{m'} + \phi_m^- (1 - T_-^2) \Delta V_m \right] \qquad (5.3)
$$

where we shorten the hyperbolic tangent with $T_\pm = \tanh(\beta(\pm h \pm \gamma m + \gamma/N))$ and the fraction of agents with a given choice as $\phi_m^\pm = (1 \mp m)/2$. For any m the $\Delta V_\cdot$ are strictly positive by Proposition 12 and so are $\phi_\cdot^\pm$. Since $1 - T_\cdot^2$ are a unimodal functions then by Lemma 5 the maximum of the r.h.s. of (5.3) happens somewhere in the interval $h^* \in (-\gamma m + \gamma/N, -\gamma m - \gamma/N)$. This is, in an absolute value sense, the highest that $\bar{h}$ can get. Hence, we show that at $m = 0$, this maximum is not attained if the discount factor is large enough. In particular, at $m = 0$ the inequality

$$
-h^* < \frac{\delta\beta}{4} \left(1 - T_+^2 \right) \phi_m^+ \Delta V_{m'} + (1 - T_-^2) \phi_m^- \Delta V_m \big] \,,
$$

holds. Hence, the optimal policy is smaller than h^*.

To show this write out the optimality equation for the short term planner at $m = 0$ and evaluated at $h^* = -\gamma m - \gamma/N$ noting that $T_+ = 0$ and $T_- = tanh(2\gamma/N) \approx O(2\gamma/N)$. Under our maintained assumption $\eta = -1$

$$\begin{aligned}
\gamma/N &< \frac{\delta\beta}{4}\left[\Delta V^0_{m'} + (1 - O(2\gamma))\Delta V^0_m\right] \\
&= \frac{\delta\beta}{4}\left[\frac{4}{N}(m'+1) - \frac{1}{N^2} + (1 - O(2\gamma))\left(\frac{4}{N}(m+1) - \frac{1}{N^2}\right)\right] \\
&= \frac{\delta\beta}{4}\left[\frac{8}{N} + \frac{6}{N^2} - O(2\gamma)\frac{4}{N} + O(2\gamma)\frac{1}{N^2}\right] \\
&= \delta\beta\left[\frac{2}{N} + O\left(\frac{1}{N^2}\right)\right]
\end{aligned} \tag{5.18}$$

Picking the appropriate δ this inequality holds, furthermore having picked the worst case h^* implies that the intersection is unique for the myopic planner. In turn, by lemma 4 the claim follows $\square$.

Chapter 6

Discussion

This thesis presented a reduced-form model of herding where boundedly rational — that is myopic, inhert and error-prone — agents making binary choices attempt to dynamically coordinate with each other and with the policy of an external planner who has an exogenous preference for one of the two choices.

One of the major question in game theory is how *equilibrium selection* happens. This work tackles the topic of equilibrium selection from a different viewpoint, asking instead: how do transition across equilibria happen? And what is the optimal policy to drive such transition?

The crucial challenge is to endow agents with a credible boundedly rational dynamic. The assumptions usually made in evolutionary game theory allow us to work with simple stochastic agents whose probability of taking a given action depends on how convenient it is. In turn, when the model is in continuous time — and any discrete Markov chain can be easily brought to continuous time — the waiting time before an agent is called to act will be such that the more convenient it is to change one's choice the sooner agents will be called to act. Here I have chosen the so called "Metropolis" dynamics: agents are randomly selected one by one and revise their choice. This seems the most natural choice, but other are possible, in fact almost any selection probability can be chosen[1] and still obtain the same stationary distributions derived here. At least in this simple model this only alters the time before the Markov chain attains its equilibrium distribution, but not the shape of the distribution itself. But as argued below in the discussion about the model with more than two choices, this might not always be the case.

The cost to pay is that rational expectations have to be left behind. I believe it not to be too steep a price. First, because rational expectations are, in some sense,

[1] For a discussion about this see Newman and Barkema (1999).

responsible for the presence of multiple equilibria. Secondly, because evolutionary game theory has shown it is possible to trade "rationality of the species for the rationality of the individual" and obtain similar results, at least to some degree. Thirdly, because I think that adding rational expectation in this model would not really change the results , but only alter convergence times of the underlying Markov chain as discussed below. This might even be considered a gain, if one thinks of rational expectations as too simplistic a modeling device, that only allow for equilibrium analysis and not for the study of dynamics.

The other conceptual advancement with respect to similar models in the field is to employ Markov chain not just as a tool to describe equilibria outcome, but to describe the dynamics itself. It is reasonable to ask: why not use the language of dynamic systems instead? The reason is that Markov chains, and more generally, stochastic dynamics, are in fact easier to deal with, especially if one wants to be able to carefully pick agents microfoundations. When the models enjoy the high degree of symmetry prescribed by Theorem 5 lumping allows for further simplification to describe aggregate behavior. Lumping reveals also how careful one must be in drawing conclusions based on the stationary distribution: Kandori et al. (2008) claim that the equilibrium distribution spends most time on the highest probable states, but they are not accounting for the fact that *highly probable states might be less frequent*. According to them our model should spend most of its time in one of the two highly coordinated states, but the presence of long-lived equilibria away from such configurations contradicts them. This is because the random selection mechanism *acts opposite* agents' utility , to which physicist refer as the fight between energy and entropy. Indeed, for high enough rationality parameter the stationary distribution peaks over the more coordinated states.

Lastly, I have shown how to analyze the optimal policy when a planner tries to steer the population behavior by applying a tax that enters directly into the utility function.

When the planner is able to observe the population only with a significant lag, as in Chapter 4, it is as if he only interacted with the stationary distribution of the Markov chains. In turn, there's a unique optimal policy value which. The first order condition shows that this condition crucially depends on the population size N, with larger population eliciting a stronger response from the planner. When the rationality parameter β is larger the planner policy will also be larger. Indeed, these two parameters act on the long run behavior of agents in a similar way, by increasing the pull of attracting regions, either because of stronger peer pressure or because of lower propensity of agents to revise their choice. Intuitively, the planner

policy depends on the skewness and variance of the stationary distribution: larger variance benefits the planner as it gives the population a higher chance to cross from one long-lived equilibria to another. The effect of skewness depends instead on whether the skewness favors the planner, making one of the two attracting regions bigger or smaller. All of these are elements that are completely exogenous to the planner. But one that is assumed to be, but need not to, is policy marginal costs. Under the assumption in this chapter, marginal costs inversely affect the amount of policy. Indeed, when the costs are low enough, I have shown that the optimal policy is guaranteed to "delete" the unfavorable long-lived equilibria. This suggests that in real-world situation, when strong herding forces are present, a lot of intervention might be necessary to steer a population toward a desired outcome, and the real bottleneck that prevents it is the marginal costs. This crucially depend on the fact that times spent in the long-lived equilibria are significantly "long", as they are always of order e^N. Until the long-live equilibria are present small adjustments to how strongly attracting these regions are only yield small benefits. The last observation that is in order about 4 is that it shows that not all crowds are the same: depending on the strenght of the coordination motiv γ, population sharing the same amount of rationality might require radically different intervention, since the optimal policy is not always monotone in β.

If the planner is able to monitor closely all decisions by agents, he can influence agents transition probability directly. As a result, the optimal policy is now state dependent and changes according to the average choice at any given time. Apart from the case where coordination is low, the policy is never monotonic in the state. The point where the highest amount of policy is applied is shown to be somewhere when more than half of the population is against the planner. After that point the optimal policy either tapers off, or suddenly jumps to a lower level. This suggests that there is some critical window for the planner to intervene: once the population as reached a certain threshold, which corresponds to entering a strongly attracting region, there is little gain to intervention. Depending on the parameter this effect can be so strong to lead to a "give up" region where the planner policy is extremely low. Again, marginal costs represent a crucial parameter and are able to enforce monotonicity of the optimal policy.

It should be noted that throught the thesis the concept of "time spent" is left purposely vague and up to interpretation. The choice by the agents could take up extremely short time span, a few instants, the time to make up one own mind; a slighlty longer one, such as the distance between one cigarette and the next; or the time at which buy and sell decisions are made in the stock market.

In the remainder of this chapter I speculate on a few possible extensions to the model presented here and on how results presented could change.

6.1 Finite size scaling

Throughout this thesis population size was some fixed, albeit possibly large, number N. What happens in the limit of $N \to \infty$? One consequence is that all probability of the stationary measure will concentrate on the state with largest probability. If there are multiple states sharing the same probability, probability will be equally split between those. This suggests that the presence of even the slightest asymmetry leads to the disappearence of long-lived equilibria: regardless of how much people care about coordination, the pull due to peer pressure becomes infinite as N grows large. In turn, the dynamic behavior of the system might change drastically and so will the policy.

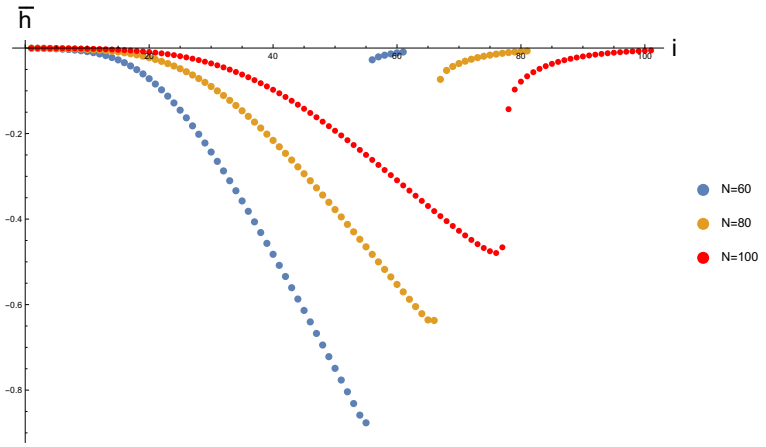


Figure 6.1: Optimal policy $\bar{h}$ under the assumptions of Ch. 5 for different values of the population size, showing the sudden "jump" policy undergoes and how it changes as N grows larger. Numerical exploration indicates that the jump is always present for finite N, albeit becoming smaller in absolute value. Horizontal axis counts the number of agents choosing $x_i = 1$.

Figure 6.1 shows the finite size scaling of the optimal policy shown in Chapter 5. The optimal policy is shown in the regime where the rationality parameter β and the strenght of the coordination motiv γ are high enough to make the optimal policy "jump" suddenly to a lower level past a certain threshold. It is clear that the optimal policy will in general be lower, since the presence of more agents reduces the marginal gains for the planner. It remains to be shwon whether the "jump" completely disappears in the limit $N \to \infty$.

6.2 Multiple choices

It is relatively easy to extend the model to allow for multiple choices. Agents called to action would then receive a proposal to switch from their current choice to a different one picked at random. With q possible choices, the predominant choice in the population can than be represented by a vector in the complex plane:

$$m = \sum_{p=1}^{q} e^{2\pi i \frac{p}{q}} n_p,$$

this funtion weights each of q directions in the complex plane by n_p, the fraction of agents currently choosing the $p-th$ choice.

The mathematics of such a system become much more involved. In particular, the lumped version of the system looses the Markov property. There are reasons to believe the same type of long lived equilibria would appear in this case and that the policy would show a similar behavior. Indeed, Slowik (2012) shows the potential of such a system with $q = 3$, which is reproduced with some editing in Figure 6.2.

Figure 6.2: Plot of the potential as a function of the average choice $m = \sum_{p=1}^{q} e^{2\pi i \frac{p}{q}} n_p$ mapped to the complex plane, when agents can choose between $q = 3$ different choices. The three deepest potential well are located along the vectors that represent the population being fully coorinated on one of the choices. The central one represent a long-lived disordered state with no clear majority, due to the particular probability rates chosen. Figure is edited from Slowik (2012) Ph.D. thesis on the Potts model.

The figure highlights the presence of four potential wells, three are the attractors due to different choice, but one more emerges. The reason for this is that the model behind the figure is generated with metropolis rates: agents are selected at random and proposed a random choice out of the q available to them. The central potential well is therefore present in a situation of high-disorder when there isn't any strong majority, in which case agents might be switching randomly for quite a while, befor finally transitioning to a coordinated state. A more natural option that might solve this is to use "heat bath" dynamics, which offers agents to switch their

current choice with the one they are more likely[2] to switch to.

6.3 Absorbing endstates

One could imagine that once the totality of the population agreed on one choice the game terminates. For example, reaching -1 could be seen as triggering a revolution and 1 as deciding in favor of mantaining the status quo, at which point the coordination game would end. This could be imposed on the optimization problem in Chapter 4 by introducing absorbing boundaries over the lumped state space.

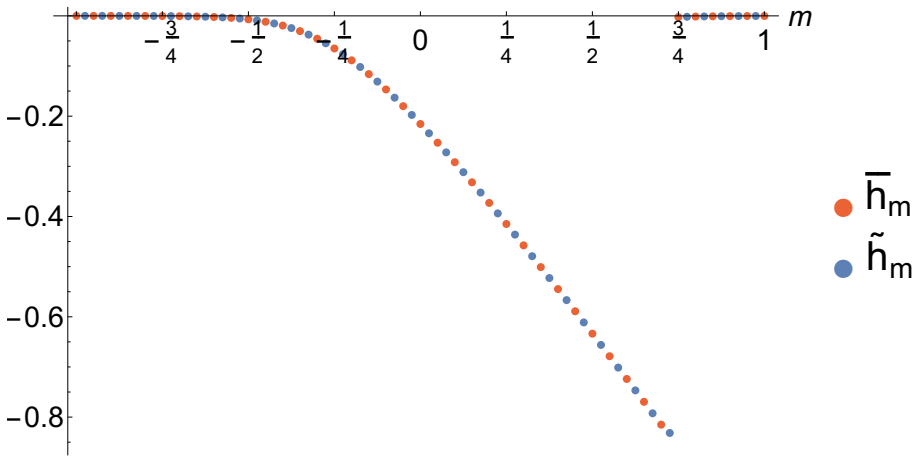


Figure 6.3: A plot of the optimal policy under the assumption of frequent sampling from the planner exposed in Chapter 5. The rationality parameter β is equal to 1.3 and the coordination motive γ is high at 1. The two plots are interleaved, showing they are the same regardless of whether the boundaries are absorbing (blue) or not (orange).

Numerical exploration suggests that absorbing boundaries do not alter the optimal policy. This is somewhat surprising, given that the first order condition in Eq. (5.3) sees a direct changes in the last two states and an indirect one through changes of the value function terms appearing in it.This suggests that the model might be employed to investigate optimal policies even for transitory phenomena. One possible explanation is that the boundaries are very rarely visited, and therefore they have a very minimal impact on the life and overall behavior of the chain. Hence, what really matters from the point of view of the planner is always the short to medium-term behavior of the chain.

[2]Contrast this with Metropolis dynamics which offer agents to switch to a random choice

6.4 Agents with higher order expectations

One criticism that might be levied against the model is that agents are myopic and don't give any amount of consideration to how their choices influence their neighbors future play. It is not obvious to say what changes might occur. On the one hand, a similar model with belief is presented in Durlauf (1996), shows that in the limit of large population and under rational expectations, the equilibrium average choice would corrispond to the bottoms of the wells of the potential function which describes our stationary distributions. On the other, forming beliefs, for example as described in Golub and Morris (2017), would alter convergence rates of the Markov chain toward its distribution. The discussion in the section above shows that the relevant part of the chain's life might be in fact the short and medium term[3].

[3]Otherwise we would expect that absorbing states that are eventually reached to influence the value function and return a different policy. The fact that the discount factor δ is always equal to 0.9 and "high" in our calculation reinforces this consideration.

Bibliography

Akerlof, R., Holden, R., Akerlof, G., Antras, P., Dixon, R., Fudenberg, D., Gibbons, B., Golub, B., Goyal, S., Li, H., Rayo, L., Stein, J., and Zingales, L. (2019). Capital Assembly. Technical report.

Akerlof, R., Holden, R., and Rayo, L. (2018). Network Externalities and Market Dominance *. Technical report.

Anderson, S. P., Palma, A. D., and Thisse, J. F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Averintsev, M. (1970). On a method of describing complete parameter fields. *Problemy Peredaci Informatsii*.

Baxter, R. J. (2016). *Exactly solved models in statistical mechanics*. Elsevier.

Blume, L. E. (1993). The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior*, 5(3):387–424.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear Social Interactions Models. *Journal of Political Economy*, 123(2):444–496.

Bouchaud, J.-P. (2013). Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges. *J Stat Phys*, 151:567–606.

Bovier, A. and den Hollander, F. (2015). *Metastability: A potential theoretic approach*, volume 351 of *Grundlehren der mathematischen Wissenschaften*. Springer International Publishing, Cham.

Brock, W. A. and Durlauf, S. N. (2001). Discrete Choice with Social Interactions. Technical report.

Burke, M. A., Young, H. P., Bisin, A., Benhabib, J., and Amsterdam, M. J. (2011). Chapter 8 - Social Norms. *Handbook of Social Economics*, 1:311–338.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646.

Clark, A. E. and Oswald, A. J. (1998). Comparison-concave utility and following behaviour in social and economic settings. *Journal of Public Economics*, 70:133–155.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (1992). Social Norms, Savings Behavior, and Growth. Technical Report 6.

Devenow, A. and Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, 40:603–615.

Durlauf, S. (1996). Statistical Mechanics Approaches to Socioeconomic Behavior. *NBER Technical Working Paper Series*, 203.

Ellison, G. (1993). Learning, Local Interaction, and Coordination. *Econometrica*, 61(5):1047.

Ellison, G. (2000). Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Review of Economic Studies*, 67(1):17–45.

Feddersen, T. and Pesendorfer, W. (1997). Voting Behavior and Information Aggregation in Elections With Private Information. Technical Report 5.

Friedman, D. (1998). On economic applications of evolutionary game theory. *Journal of Evolutionary Economics*, 8(1):15–43.

Fundenberg, D. and Maskin, E. (1990). Evolution and Cooperation in Noisy Repeated Games. Technical Report 2.

Golub, B. and Morris, S. (2017). Expectations, Networks, and Conventions. *SSRN Electronic Journal*.

Harsanyi, J. C. and Selten, R. (1988). A General Theory of Equilibrium Selection in Games. *MIT Press Books*, 1.

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, Mutation, and Long Run Equilibria in Games. *Econometrica*, 61(1):29–56.

Kandori, M. and Rob, R. (1995). Evolution of Equilibria in the Long Run: A General Theory and Applications. *Journal of Economic Theory*, 65:383–414.

Kandori, M., Serrano, R., and Volij, O. (2008). Decentralized trade, random utility and the evolution of social welfare. *Journal of Economic Theory*, 140(1):328–338.

Kawagoe, T., Matsubae, T., Takizawa, H., and Takizawa, H. (2018). Quantal response equilibria in a generalized Volunteer's Dilemma and step-level public goods games with binary decision. *Evolutionary and Institutional Economics Review*, 15(1):11–23.

Kemeny, J. and Snell, J. (1960). *Finite Markov Chains*. The University Series in Undergraduate Mathematics. D. Van Nostrand Co., Inc., Priceton-Toronto-London-New York.

Kobayashi, H., Mark, B. L., and Turin, W. (2011). *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press.

Kochmański, M., Paszkiewicz, T., and Wolski, S. (2013). Curie-Weiss magnet - A simple model of phase transition. *European Journal of Physics*, 34(6):1555–1573.

Krugman, P. (1998). What's new about the new economic geography? *Oxford Review of Economic Policy*, 14(2):7–17.

Liggett, T. M. (1985). *Interacting Particle Systems*. Springer New York.

Lucas, R. (1976). Economic Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, pages 19–46.

Luce, R. and Suppes, P. (1965). Preference, Utility, and Subjective Probability. In Luce, R., Bush, R., and Galanter, E., editors, *Handbook of Mathematical Psychology III*, pages 249–410. Wiley, New York.

Mailath and J., G. (1992). Introduction: Symposium on evolutionary game theory. *Journal of Economic Theory*, 57(2):259–277.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. and McFadden, D., editors, *Structural analysis of discrete data with econometric applications*, chapter Econometri, pages 198–272. MIT Press, Cambridge.

McFadden, D. L. (1984). Chapter 24 Econometric analysis of qualitative response models. In *Handbook of Econometrics*, volume 2, pages 1395–1457. Elsevier.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior*, 10(1):6–38.

Newman, M. E. J. M. E. J. and Barkema, G. T. (1999). *Monte Carlo methods in statistical physics*. Clarendon Press.

Norris, J. R. (1998). *Markov chains*. Cambridge University Press.

Ross, S. M. (1983). *Introduction to stochastic dynamic programming*. Academic Press.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2):143–186.

Schelling, T. C. (1973). Hockey Helmets, Concealed Weapons, and Daylight Saving. *Journal of Conflict Resolution*, 17(3):381–428.

Slowik, M. (2012). Contributions to te Potential Theoretic Approach to Metastability. *PhD Thesis*.

Smith, J. M. and Price, G. R. (1973). The Logic of Animal Conflict. *Nature*, 246(5427):15–18.

Spitzer, F. (1971). Markov Random Fields and Gibbs Ensembles. *The American Mathematical Monthly*, 78(2):142–154.

Volij, O., Ben-Shoham, A., and Serrano, R. (2004). The Evolution of Exchange. *Staff General Research Papers Archive*.

Young, H. P. (1993). The Evolution of Conventions. *Econometrica*, 61(1):57–84.