

Essays on the Evolution of Social Co-ordination
and Bounded Rationality

Tom Quilter

PhD in Economics

The University of Edinburgh

2009

Abstract

Many evolutionary game theory papers have obtained their results when the bounded rationality which creates change vanishes. In our first chapter we consider whether such results are actually a good reflection of a population whose bounded rationality is small yet persistent. Our model consists of a two type population with three stable equilibria. Firstly we find that results from the standard vanishing noise approach can be very different from those obtained when noise is small but constant. Secondly when the results differ the small and persistent noise approach selects an equilibrium with a co-existence of conventions.

Our second chapter generalises the model of our first chapter to a population of many player types and several stable equilibria. Firstly we produce the characteristics of the long run equilibria under vanishing noise analysis. Secondly we find that the introduction of a small neutral group into a divided society can produce a welfare improving switch in the long run equilibrium towards social co-ordination.

Our third chapter combines the model of the second chapter with the message of the first. We show numerically that the long run location of a heterogenous population with extremely low levels of bounded rationality can be completely different to the equilibria selected through vanishing noise analysis. We also show that such an event is not a rare occurrence and find that over a third of populations are misrepresented by stochastic stability.

Our final chapter conducts a review of the literature on social threshold models. We give a thorough description of each paper and discuss the main assumptions that drive the key results.

Contents

1	Introduction	4
2	Noise Matters in Heterogenous Populations	8
2.1	Introduction	9
2.2	The Model	14
2.2.1	$\rho = 1$ and Pairwise Matching	14
2.2.2	$\rho < 1$: A Public Smoking Example	16
2.2.3	Players and the Stochastic Dynamics	17
2.2.4	Welfare	21
2.3	An Illustration of the Results	22
2.4	Analysis	24
2.5	Calculations	34
2.5.1	Minority groups	37
2.6	Conclusion	38
2.7	Appendix	40
3	Three's a Crowd	51
3.1	Introduction	52
3.2	The Model	55
3.2.1	The Dynamics of a Heterogenous Population	56
3.3	Stochastic Stability	59
3.4	The Influence of a Neutral Group in a Three Type Population	62
3.4.1	The Analysis	64
3.4.2	Pairwise Matching	66
3.4.3	Equilibrium Selection Results	68

3.4.4	The Effect of the Introduction of a Neutral Group	71
3.5	Conclusion	73
3.6	Appendix	75
4	The Limitations of Stochastic Stability in Heterogenous Populations	84
4.1	Introduction	85
4.2	The Model	87
4.2.1	The Dynamics of a Heterogenous Population	88
4.3	A Positive Level of Noise	90
4.4	An Introduction to the Results	93
4.5	The Limitations of Stochastic Stability	96
4.6	The Prevalence of Stochastic Stability	
Misrepresentation		99
4.7	Discussion of the Results and Dynamics	101
4.8	Conclusion	105
5	Models of Evolution and Social Thresholds	108
5.1	An Introduction to Social Threshold Models	109
5.2	Dynamic Models of Segregation	
–	Thomas.C.Schelling 1971	113
5.3	Threshold Models of Collective Behavior	
–	M.Granovetter 1978	116
5.4	Chains of Co-operative Action: Threshold Effects in Collective	
Action - M.W.Macy 1991		120

5.5	Threshold Models of Diversity: Chinese Restaurants, Residential Segregation and the Spiral of Silence - Granovetter and Song 1998	125
5.6	Universal Behavior in a Generalized Model of Contagion - Dodds and Watts 2004	130
5.7	Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning - P.Young 2008	134
5.8	Social network thresholds in the diffusion of innovations - Va- lente 1996	138
5.9	Conclusion	141

1 Introduction

Many interesting games have more than one Nash equilibrium, and so it can be difficult to determine which equilibrium one would expect to see when the game is played by real people. Evolutionary game theory has made a particularly successful attempt in attacking the problem. By allowing players to be boundedly rational and change their mind at any point, through experimentation, lack of concern or irrational behaviour, a population will not get stuck doing the same thing for a long time. And so a population of boundedly rational players will fluctuate between different equilibria.

However the aim is not to create a dynamic population, but to select a unique equilibria which we would expect the population to be in. The most popular method of achieving this is to allow the bounded rationality of each player to vanish, in that the probability of a player changing their strategy tends to zero. One can think of a boundedly rational population as a box of warm particles, which are moving around and bouncing between two attractors. The method of vanishing bounded rationality can then be considered similar to cooling the box. The particles become slower and slower, and eventually congregate by one attractor. This method of vanishing noise has been very successful as it pin points exactly one equilibria for the population to be in. Furthermore it is attractive due to its relative ease in obtaining analytical results for researchers. The justification for obtaining results using vanishing noise is that they will accurately reflect a population that has small, but realistic amount of noise. This seems very reasonable, and as such most research has focused its attention on obtaining their results as noise vanishes. However, the main aim of our first chapter is to examine whether vanishing noise results are actually a good re-

flection of a population whose bounded rationality is small, but non-vanishing. To achieve this we consider a method which allows a persistent level of bounded rationality to remain for each player. A population subject to constant noise is always in motion. However, for any particular population size, it is likely that a population spends more time playing one type of strategy than another. Indeed, as we let the population size increase without limit we find a single neighbourhood where the process spends all of its time is selected.

The chapter is designed with an intention to compare the results of vanishing noise with those obtained under positive noise with an increasing population. This is considered within the context of a model of a heterogeneous population, and with it a conflict of interest. The model has two types of players within a population playing a coordination game, who differ in their preferred choice of two strategies. This heterogeneity creates three separate equilibria. Two where every member of population plays the same strategy, and a third where both strategies are played in the population.

In our main result we discover that there is a range of situations where a large population with a only a small amount of non-vanishing noise will be in a very different place to where vanishing noise analysis tells us. The two methods yielding completely different results. Furthermore, in our second result we find that when a discrepancy between the two methods exists, it is the non-vanishing noise approach that selects the co-existence equilibrium, vanishing noise selecting one of the monomorphic equilibria instead.

And so I feel that the results of the first chapter are a strong indication that vanishing noise results can be quite wrong, and that there should be much more consideration of truly boundedly rational populations. Furthermore, the prominence of co-existence equilibria may have been underestimated through

the popularity of vanishing noise methodology.

The second chapter seeks to expand on the first chapter by introducing a general level of heterogeneity into the model, attempting to continue to link the work of Kandori, Mailath, Rob and Young with the general binary-action games of the kind first introduced by Schelling. We investigate a population consisting of many different types of player who vary in their preferences for two strategies in a co-ordination game. The general level of heterogeneity can create many equilibria. One possible application among many could be a key vote in a society between two opposing issues, such as the U.S presidential election. Here voters often prefer others to agree with them and there invariably exists a large range of different views within the population. There are many stable voting proportions with the final ratio being of great importance.

We first look to addresses which of the many equilibria the population will spend most of its time in the long run. With the increased heterogeneity positive noise analysis is complex and we obtain our results as the noise level vanishes. For any general degree of heterogeneity, we are able to determine the precise mathematical characteristics of the long term location of the population.

We continue the chapter by applying this result to investigate the influence a neutral group may have upon a divided society. We show that in some cases the introduction of only a small amount of neutral agents can upset a co-existence equilibria and sway society to full agreement.

Our third chapter links the first two chapters together and in particular seeks to continue our work on the limitations of vanishing noise from chapter one. By taking our general heterogenous model from chapter two and selecting specific populations we are able to produce exact numerical results on the long run

location of the society for different positive noise levels.

We show that increasing the heterogeneity from the two type model of our first chapter extenuates the chapter's main result. Indeed we find some populations with individual experimentation rates as small as one in a million periods can in fact be located in a completely different neighbourhood in the long run to the stochastic stability equilibrium. And so we show that populations under extremely small noise levels can be located in very different neighborhoods to where vanishing noise would suggest. We also find that over half of populations with more than 4 player types are misrepresented by stochastic stability.

Our final chapter conducts a review of the social threshold model literature. Such models are driven by the assumption that people are strongly influenced by other members of society, and when enough people take up an action others will may also be persuaded to join in. The review begins with Thomas Schelling's ground breaking paper on social segregation and continues to review a variety of research papers stemming from Schelling's original idea. We give a thorough description of the models of each paper and provide discussion of the assumptions that drive the main results. We also note that some results may not prevail if they were subjected to small perturbations.

2 Noise Matters in Heterogenous Populations

Abstract

The concept of boundedly rational agents in evolutionary game theory has succeeded in producing clear results when traditional methodology was failing. However the majority of such papers have obtained their results when this bounded rationality itself vanishes. This paper considers whether such results are actually a good reflection of a population whose bounded rationality is *small*, but *non-vanishing*. We also look at a heterogenous population who play a co-ordination game but have conflicting interests, and investigate the stability of an equilibria where two strategies co-exist together. Firstly, I find that results using the standard vanishing noise approach can be *very different* from those obtained when noise is small but persistent. Secondly, when the results differ it is the non-vanishing noise approach which selects the co-existence equilibria. As recent economic and psychology studies highlight the irrationality of their human subjects, this paper seeks to further demonstrate that the literature needs to concentrate more on the analysis of truly noisy populations. Keywords: Non-vanishing noise, equilibrium selection, strategy co-existence.

2.1 Introduction

Nash Equilibrium has been the corner stone of game theory, however the existence of multiple Nash equilibria in even the simplest of games has proved a stubborn obstacle for theorists. When a population of rational players are in one of the Nash equilibria, the population is stuck there.¹ This equilibria could be the least efficient. And the only factor determining the equilibria selected are the preliminary beliefs of the population. Evolutionary game theory has led the quest to find more appealing solutions. The ground-breaking papers of Kandori, Mailath and Rob(1993) and Young(1993)² produced a significant insight. By introducing boundedly rational agents who occasionally make mistakes, a population now had the potential to move between multiple equilibria. This persistent noise gives the process life, allowing for the investigation of which equilibria the population is more likely to be near, independent of the initial conditions.³

However when analysing a population subject to persistent noise, which by its nature is continually moving between states,⁴ it is difficult to obtain clear results of whereabouts it will be in the long run. KMRV overcame this issue by producing all their results from analysis as the noise level decreases to 0. Here, for a population of any size, a single state is solely selected in the long run as noise vanishes.⁵

¹By definition, no-one has an incentive to deviate.

²KMRV from here on.

³The introduction of boundedly rational agents also served as a step to address the criticisms of analysis with hyper-rational players, players of god like intelligence and endless time to use it.

⁴Each state specifies a different combination of who is playing each strategy.

⁵The trend of vanishing noise analysis has continued. Early vanishing noise papers including Ellison(1993), Samuelson(1994), Begin and Lipman(1996), Fernando and Vega-Redondo(96) and Ellison(2000) and more recently Kolstad(2003), Myatt and Wallace(2003), Norman(2003a) and Hojman(2004).

In this paper we also consider an alternative method, in which we allow a constant level of noise, but let the population size increase without limit. With non-vanishing noise a single state can never be selected as the process is always in motion. Nevertheless, for any population size, it is likely that process spends more time in one neighborhood of the state space than another. Indeed, as we let the population size increase without limit, we find a single neighborhood where the process spends all of its time is selected.

I feel that as we are dealing with bounded rationality, the second method makes more intuitive sense. As vanishing noise results require the source of the dynamics to disappear, it seems that such results are only justified if they reflect those that would be obtained under small non-vanishing noise, with boundedly rational agents who actually do make mistakes occasionally.

And so the primary aim of the paper is to assess whether the two methods agree on the long run location of the process. Specifically, in a large population, is the state selected under vanishing noise always within the neighborhood selected under small non-vanishing noise?

Surprisingly, in our main result we find that in some circumstances a large population with a only a small amount of non-vanishing noise⁶ will never be where vanishing noise analysis tells us, the two methods yielding completely different results. The single state chosen under vanishing noise can be very far from the neighborhood selected under non-vanishing noise and an increasingly large population.

And so here we see vanishing noise analysis can present a very misleading portrayal of an actual boundedly rational society. Consequently I believe

⁶One mistake every one hundred periods in the example of section 3.

that there exists a dangerous trend in the literature to conduct vanishing noise analysis alone, without any consideration of how significantly their results may change under non-vanishing noise. Often there are no simulations or calculations in the papers.

To explain the second aim of the paper and our choice of model, let us consider the mobile phone market in the UK. The existence of high call charges to networks other than your own entails that each individual prefers the whole population to be on their own network. Therefore the most efficient market set-up would be one where just one network exists. And indeed, in the homogenous population of KMRV we find that the population is only stable when the entire population plays one strategy. Yet interestingly, when we look at the actual mobile phone market we continually observe many networks co-existing together. And there are other important examples of such strategy co-existence. Most notably, we often see many different political and religious beliefs existing within a population,⁷ and this lack of co-ordination can sometimes produce severe inefficiency. On a smaller scale, different members of a town will often choose to invest in different public goods. Even towns which follow two sports teams could well be better off with everyone supporting just one.⁸ Although there are probably several reasons for strategy co-existence, this paper wishes to explain such observations by allowing different people to like different things. And so we may see one section of society playing the strategy they prefer, while the rest of the population play a different strategy which they favor. Therefore I choose a model which has 2 types of players within a population, differing in their preferences for two strat-

⁷As an individual you often prefer more people in the population agreeing with your own beliefs, than less.

⁸Some UK populations are known as either a football town, or a rugby town.

egy choices. This increased heterogeneity creates a third equilibrium, a co-existence equilibrium, where both strategies are played in the population.

We find under both methods that the co-existence equilibria can easily be the long-run location of the process. Furthermore, we find that when a discrepancy between the two methods exists, it is the non-vanishing noise approach that selects the co-existence equilibrium, vanishing noise selecting one of the monomorphic equilibria instead. And so by considering a population that is both heterogenous and boundedly rational, we reveal that observing several strategies together is possible, and indeed likely. Therefore this type of equilibrium most likely plays a much larger role in more realistic populations than homogenous populations under vanishing noise would suggest.

Analysis with non-vanishing noise is not unique to this paper. For example Benaim and Weibull [2003a,b] also keep noise constant while taking population size to infinity, as do Binmore and Samuelson(1997). Myatt and Wallace(1998), and Beggs(2002) also devote some attention to the concept. The most similar paper is probably that of Sandholm(2005), which also looks at constant noise while taking population size to infinity, showing that in a homogenous population there can exist some type of game where it is possible for constant noise results to differ from those of vanishing noise. The quantal response literature (McKelvey and Palfrey, 1995) has also given us the best experimental evidence that positive noise does matter. Separately, the co-existence of strategies exist in Kolstad(2003) and Anwar(1999) to name two. In Norman(2003) the introduction of switching costs creates many stable points of co-existence, although the vanishing noise analysis employed showed that in the long

run no time would be spent in these states.

Here we look to investigate a model with heterogeneity, co-existence equilibria and positive noise, and the paper reads as follows. In Section 2 we present the model of a boundedly rational population with players of two conflicting types who have a choice of two strategies, with similarities to Kolstad's(2003) fluid interaction. They play a game of co-ordination as in any period an agent's payoff is monotonically increasing in the number of other agents playing her chosen strategy.

In Section 3, we give a quick and clear illustration of our main results.

In Section 4, the framework of the analysis is set out. The selection results pertaining to when the population spends all its time near the co-existence equilibrium are obtained using the usual vanishing noise approach, then selection results are instead found with small non-vanishing noise and the population size being allowed to increase without limit. The results are then compared, producing the main result.

Section 4 illustrates a sample of real calculations of boundedly rational populations, showing that the larger the rate of individual mistakes, the more likely the two strategies will co-exist. We also take a look at the survival of minorities groups. Section 5 concludes.

2.2 The Model

Let a single population of N players consist of two types of players, type 1 denoted T_1 and type 2, T_2 . The game is essentially one of co-ordination as in any period the more players in the population playing an agent's current strategy, the higher is that agent's payoff. However, there are two different types of players who receive different payoffs each period.

The payoff in any period t for a T_1 agent playing strategy s_i , $\pi_1^{s_i}$, is given by

$$\pi_1^{s_1} = \beta_1(z(t) - 1)^\rho \quad (1)$$

$$\pi_1^{s_2} = \gamma_1(N - z(t) - 1)^\rho \quad (2)$$

where $z(t)$ represents the number of agents playing s_1 in period t and $\rho \in \mathbb{R}^{+9}$. $\beta_1 > \gamma_1$ indicates that all T_1 agents have the same preference to co-ordinate on strategy 1 rather than 2.

The essential difference between T_2 and T_1 players is that T_2 agents prefer the population to co-ordinate by playing s_2 rather than s_1 , while T_1 agents have the opposite preference.

And so we have it that the payoff for a T_2 agent playing strategy s_i , $\pi_2^{s_i}$, is given by

$$\pi_2^{s_1} = \beta_2(z(t) - 1)^\rho \quad (3)$$

$$\pi_2^{s_2} = \gamma_2(N - z(t) - 1)^\rho \quad (4)$$

where $\beta_2 < \gamma_2$.

2.2.1 $\rho = 1$ and Pairwise Matching

Here we see that a special case of the model is the familiar idea of pairwise patching, that in each period an agent has an equal chance of playing a

⁹For most applications we would have $\rho \in (0, 1]$ but we leave $\rho > 1$ open for generality.

stage game with any other agent in the population.

While demonstrating this, let us consider an example. Competing cell phone companies often have far higher charges for calls to other networks than to calls to the same network, thus each call is a co-ordination game. Consider a heterogenous world where T_1 agents (person or firm) prefer the network orange over T-mobile (perhaps due to differing sms packages, etc), and T_2 agents favor T-mobile. Then stage game between the two is given by¹⁰

T_1, T_2	<i>Orange</i>	<i>T - mobile</i>
<i>Orange</i>	a, c	e, f
<i>T - mobile</i>	g, h	b, d

where $a > e$, $b > g$, $c > f$, $d > h$ indicates the co-ordination nature of the game and $a - g > b - e$ and $c - h < d - f$ reveals the different preferences of the two types. Without loss of generality e, f, g and h can be set to 0, and therefore the three stage games are

T_1, T_2	<i>Or</i>	<i>Tm</i>	T_1, T_1	<i>Or</i>	<i>Tm</i>	T_2, T_2	<i>Or</i>	<i>Tm</i>
<i>Or</i>	a, c	$0, 0$	<i>Or</i>	a, a	$0, 0$	<i>Or</i>	c, c	$0, 0$
<i>Tm</i>	$0, 0$	b, d	<i>Tm</i>	$0, 0$	b, b	<i>Tm</i>	$0, 0$	d, d

The matching process is one phone call each period to any other member of the population (equally likely). The payoff represents the cheapness of the call rate to the individual. Every period each agent will decide whether to change his network or not depending on how many people are on each network and his preferences, experimenting on occasions.

Now, by setting $\rho = 1$, $\beta_1 = \frac{a}{N-1}$, $\beta_2 = \frac{c}{N-1}$, $\gamma_1 = \frac{b}{N-1}$, and $\gamma_2 = \frac{d}{N-1}$ in equations 1 to 4, we have it that $\pi_j^{s_i}$ becomes the average expected payoff

¹⁰There are two other equally important stage games, one between two T_1 agents and another between two T_2 agents.

for a T_j agent playing strategy s_i for pairwise matching. And therefore pairwise matching is just a special case of the general model.

2.2.2 $\rho < 1$: A Public Smoking Example

For a further example let s_1 represent the choice of going to a smoking area and let s_2 represent going to a non-smoking area. Label T_1 agents as smokers, and T_2 agents as non-smokers. Set $\rho < 1$ and for interest consider smokers to be in the minority.

Consider in each period that two groups form within the population, one containing all the people who choose to congregate in the smoking area, the other containing those who choose not to. For instance we could be in a familiar office setting where during daily breaks most smoker types often congregate in a different area to non-smokers. Here the co-ordination payoffs in equations 1 to 4 could represent the value of forming and enjoying relationships with other members of the group. The more people in your group the better it is for you, but as you are unlikely to talk to everyone $\rho < 1$ indicates that the value of having 3 members in your group rather than 2 exceeds that of acquiring an extra 20th member. You do not interact with people in the other group during breaks and so gain no payoff from them.

Each day an agent decides whether to convene in the smoking or non-smoking area, depending on how many people were in which group yesterday and their preferences. Occasionally experimenting with a different strategy.

One question is how will employees take their breaks in the long run, all in the non-smoking or smoking area, or will a co-existence of the two groups prevail? Another question is whether the population's level of bounded

rationality will effect the answer.

Alternatively, one could imagine the population to be the regular members of a bar. And if a co-existence of smoking and non-smoking groups is prevailing even though it is socially inferior, then there may well be cause for a government body to step in and ban one of the strategies.

2.2.3 Players and the Stochastic Dynamics

The players chosen here are myopic in the sense that they believe the state of play will be the same as the previous period, $z(t-1)$, and so last periods play is the only factor effecting a player's decision this period. Therefore a T_1 agent's best response this period is

$$\begin{cases} s_1 & \text{if } z(t-1) > \frac{1}{1+(\frac{\beta_1}{\gamma_1})^{\frac{1}{\rho}}} N + \frac{(\frac{\beta_1}{\gamma_1})^{\frac{1}{\rho}-1}}{(\frac{\beta_1}{\gamma_1})^{\frac{1}{\rho}+1}} \equiv pN + \delta \\ s_2 & \text{Otherwise} \end{cases} \quad (5)$$

Note that $p < 0.5 \forall \beta_1 > \gamma_1$.

And similarly a T_2 agent's best response this period is¹¹.

$$\begin{cases} s_1 & \text{if } z(t-1) > \frac{1}{1+(\frac{\beta_2}{\gamma_2})^{\frac{1}{\rho}}} N + \frac{(\frac{\beta_2}{\gamma_2})^{\frac{1}{\rho}-1}}{(\frac{\beta_2}{\gamma_2})^{\frac{1}{\rho}+1}} \equiv qN + \zeta \\ s_2 & \text{Otherwise} \end{cases} \quad (6)$$

where $q > 0.5 \forall \beta_2 < \gamma_2$

We now continue by denoting N_1 as the number of T_1 agents in a given population, and N_2 as the number of T_2 agents, such that $N = N_1 + N_2$.

¹¹Note from the pairwise matching of section 2.1, with $\beta_1 = \frac{a}{N-1}, \beta_2 = \frac{c}{N-1}, \gamma_1 = \frac{b}{N-1}, \gamma_2 = \frac{d}{N-1}$, that $p = \frac{1}{1+(\frac{\beta_1}{\gamma_1})^{\frac{1}{\rho}}} = \frac{b}{a+b}$ is the mixed equilibrium of the T_1, T_2 stage game where T_2 agents play s_1 with probability p . And $q = \frac{1}{1+(\frac{\beta_2}{\gamma_2})^{\frac{1}{\rho}}} = \frac{d}{c+d}$ is the other mixed equilibrium where T_1 agents play s_1 with probability q .

Let us define the proportion of T_1 agents in the population as $\alpha = N_1/N$. In any period let $z_1(t)$ be the number of T_1 agents playing s_1 , and let $z_2(t)$ be the number of T_2 agents playing s_1 such that $z(t) = z_1(t) + z_2(t)$. As agents do not differentiate between other players types $z(t) = \{0, 1, \dots, N\}$ can be seen to define the state of the process at any time t . In each period every player is able to choose a best response to last periods state of play. There exists at least two stable points for the process, E_1 where all agents choose to play $s_1(z = 1)$ ¹² and E_2 where all choose $s_2(z = 0)$.

Furthermore, for $pN + \delta < \alpha N < qN + \zeta$, there exists a third stable point of the process E_m where $z_1 = \alpha N$ and $z_2 = 0$. Here all T_1 agent's best response is to play s_1 as they believe enough agents will join them to make it worthwhile, while all T_2 agents choose their preferred strategy s_2 . E_m is a steady state of co-existence of both strategies.¹³

We can now define the basins of attraction of the stable points of the process. Firstly let the basin of attraction of E_i be denoted by B_i . Then B_2 is defined by any state $z(t) \in \{0, \dots, [pN + \delta]_-\}$.¹⁴ At any point in B_2 all agent's best response is to play s_2 next period. Similarly, B_m is given by $z(t) \in \{[pN + \delta]_+, \dots, [qN + \zeta]_-\}$ and B_1 by $z(t) \in \{[qN + \zeta]_+, \dots, N\}$.

¹²Let $z \equiv z(t)$.

¹³I will consider cases only where $pN + \delta < \alpha N < qN + \zeta$ holds, as other cases essentially reduce to a homogenous population as in KMRV.

¹⁴ $[x]_-$ is the largest integer below x , and $[x]_+$ is the smallest integer above x , $[x]_- = [x]_+ - 1$.

The state space $\mathcal{O} = \{z_1(t) = 0, \dots, \alpha N, z_2(t) = 0, \dots, (1 - \alpha)N\}$ and the basins of attraction can therefore be illustrated by

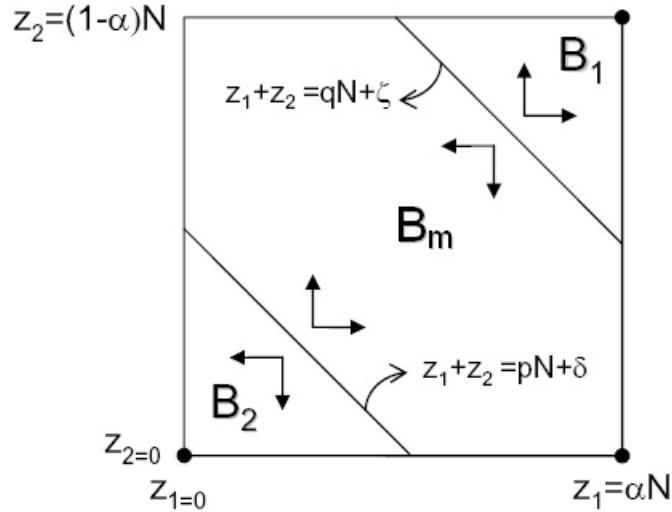


Figure 1: The State Space

Left alone the long run location of the process would depend only on which basin it was in initially. Instead, an element of bounded rationality is introduced. As in KMRY, an agent will select a strategy other than its best response with probability ε each period, I shall call such an event a mutation. Such mutations could be due to small temporary changes in circumstances for an individual. For instance, the smoking area is too cold for you one day so you go inside, your favorite football player is suspended and as a result you try a rugby game, or your mobile phone bill was unexpectedly expensive and so you change networks.¹⁵

Now via a certain number of mutations, it's possible for the process to leave its initial basin, and any other (often referred to as a basin jump).

Indeed, the process is irreducible and aperiodic as it's possible to jump

¹⁵This is my preferred interpretation. The more familiar story is that players experiment, just make mistakes or dye with probability 2ε and are replaced.

from any given state to any other state in one period, including itself, the markov chain is ergodic. Perhaps the best way to visualize the markov chain is given by the simplified state space of $z(t) = \{0, 1, \dots, N\}$ illustrated below, the larger arrows representing basin jumps, the smaller showing the flow of the basins.

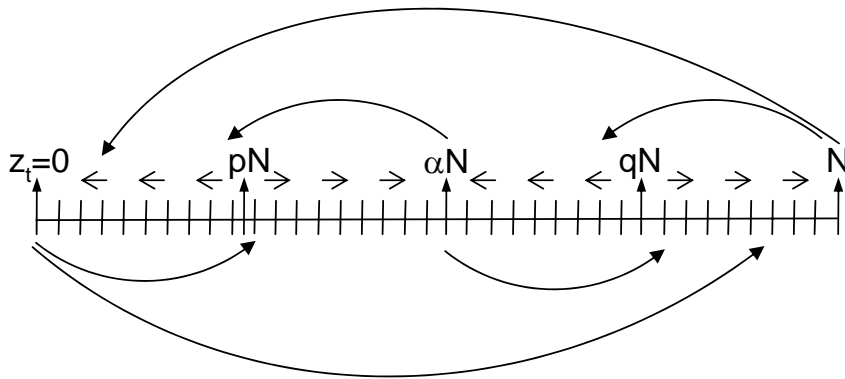


Figure 2: The Simplified State Space

And so we have a non-linear stochastic difference equation

$$z(t + 1) = B(z(t)) + q(t) - r(t)$$

given $q(t) \sim \text{Bin}(N - B(z(t)), \varepsilon)$, $r(t) \sim \text{Bin}(B(z(t)), \varepsilon)$ and where $B(z(t))$ gives $z(t + 1)$ when all agents (of both types) choose their best response to $z(t)$ last period without mutation. Thus we have a markov matrix Γ^ε with transition probabilities given by $\Gamma_{mn} = \mathbb{P}(z(t + 1) = n | z(t) = m)$.

The long run behavior of the Markov chain is given by the stationary equations $\mu^\varepsilon \Gamma^\varepsilon = \mu^\varepsilon$, the solution μ^ε is stationary for fixed Γ^ε . Indeed, for an ergodic Markov chain μ^ε will be unique and therefore independent of the initial conditions. $\mu^\varepsilon = (\mu_1, \mu_2, \dots, \mu_N)$ can be seen as the propor-

tion of time society spends in each state $z = 1, 2, \dots, N$.

Lemma 1 *The Markov chain on the finite state space $Z = \{0, \dots, N\}$ defined by Γ_{mn} is ergodic . It therefore has a unique invariant distribution, μ^ε .*

Proof. This is a standard result. For example see Grimett and Stirzaker, 2001.

2.2.4 Welfare

Before continuing let us take the opportunity to discuss social welfare in different equilibria. Welfare in the co-existence equilibrium is often lower than the two pure equilibria as here the conflict between the two groups diminishes the network effect. Even though each agent type is playing their preferred strategy, they fail to co-ordinate with a whole section of society.

As the process is always dynamic when noise is allowed to stay constant, it is difficult to make precise statements on the welfare of the society in certain neighborhoods. However, for small values of noise we can say something of the total social welfare in each of the three stable states of the population.

Lemma 2 Let $\rho = 1$ and ε be small. Then,

- a) for an equally distributed population such that $\alpha = 0.5$, if $\beta_1 \simeq \gamma_2$ then E_m is always the *worst* of the three equilibria in terms of total social welfare, and
- b) for any value of α , β_1 and γ_2 , E_m can never be the equilibria which maximises total social welfare.

Proof. Given in the Appendix. ■

2.3 An Illustration of the Results

Before we dive into the analysis, we present some calculations which demonstrate the stark difference in conclusions that vanishing and non-vanishing noise analysis can yield.

Consider a population with 50 T_1 and 50 T_2 players such that $\alpha = \frac{1}{2}$, who play the game below with the pairwise matching described in section 2.2.1.¹⁶

T_1, T_2	s_1	s_2
s_1	8, 7	4, 0
s_2	0, 4	7, 8

The graph below shows that vanishing noise analysis concludes that the population will spend *none* of its time in the basin of attraction of E_m .

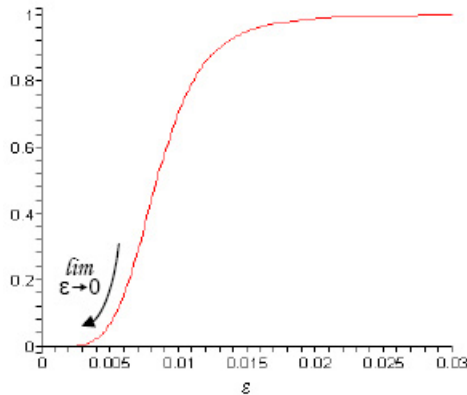


Figure 3: Time spent in the basin of E_m as noise vanishes

However, one can easily see from the graph that at an extremely small mutation rate of one in a hundred periods ($\varepsilon = 0.01$), the population in fact spends over 75% of its time in E_m 's basin of attraction.

¹⁶Note that the following results would be identical if instead the example was with $\rho = 0.5, \beta_1 = \gamma_2 = 1.63$ and $\beta_2 = \gamma_1 = 1$.

And so vanishing noise can portray a completely misleading picture of where a slightly noisy population will be in the long run.

In order to obtain clear results with positive levels of noise, we allow the population size to increase without limit. In fact, a large population with a mutation rate of one in a hundred periods will spend almost *all* of its time in the basin of E_m .¹⁷

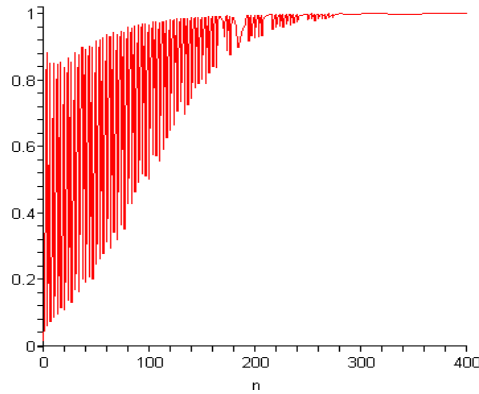


Figure 4: Time spent in the basin of E_m as the population increases, $\varepsilon = 0.01$.

And in terms of social welfare, this is the worst neighborhood for the population to be in.

¹⁷Note that in the graph below the large jumps in π_m as N increases are due to the discontinuous $[x]_+$ and $[x]_-$ functions which appear in all the markov probabilities. Also note the graph contains some non-integer values of N are not relevant, but do do little harm in illustrating the nature of the dynamics.

2.4 Analysis

As we will be dealing with positive levels of noise, the process will not converge to a single point. And so we divide the state space into three neighborhoods defined by the three basins of attraction, B_2, B_m and B_1 . Indeed, under the best reply dynamics, at any state in a particular basin of attraction all agents of a particular type have the *same* best response next period. Therefore the number of mutations required to leave a basin, and the probability of this occurring, is the *same at any state* in that basin. We can now consider just three states, $V = \{B_1, B_m, B_2\}$, and let $V(t)$ indicate which basin the population is in at time t . By defining the probability of leaving any state in basin i and entering any state in basin j by $p_{ij} = \mathbb{P}(V(t+1) = B_j | V(t) = B_i)$, we are able to simplify the whole state space into the three state markov chain below.

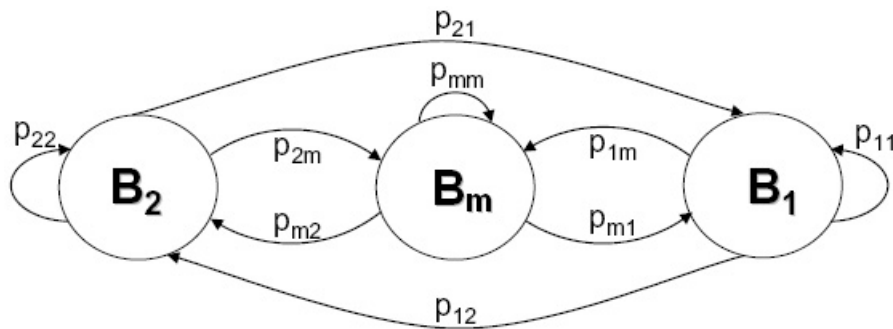


Figure 5: The Three State Markov Chain

Therefore we have a new ergodic markov chain whose long run behavior is given by the stationary equations $\pi^\varepsilon P^\varepsilon = \pi^\varepsilon$, where P^ε is the transition matrix containing the nine transition probabilities of p_{ij} , and π^ε is the unique solution for fixed P^ε . Here $\pi^\varepsilon = (\pi_1, \pi_m, \pi_2)$ can be seen as the proportion of time society spends in each neighborhood $V = B_1, B_m, B_2$.

The minimum number of mutations required to leave the basin of E_2 is $[pN]_+$, no matter which state of B_2 the process is in. Therefore the probability of escaping B_2 and entering B_m in any period is the simply the probability having between $[pN]_+$ and $[qN]_-$ mutations, and so

$$p_{2m} = \sum_{i=[pN+\delta]_+}^{[qN+\zeta]_-} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}$$

Similarly, the probability of escaping B_1 and entering B_m in any period requires between $N - [qN]_-$ and $N - [pN]_+$ mutations, and so

$$p_{1m} = \sum_{i=N-[qN+\zeta]_-}^{[N-[pN+\delta]_+]} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}$$

Escaping from E_m is a more complicated affair. Two mutations from different types effectively cancel each other out and they will have no effect on the next period's state of play (there is no change in z). Therefore the process requires a net number of mutations in one direction to make a jump.¹⁸ And so the probability of escaping B_m and entering B_1 is given by

$$p_{m1} = \sum_{j=N-[qN+\zeta]_+}^N \sum_{k=\max\{j-\alpha N, 0\}}^{\min\{j, N-i\}} \binom{\alpha N}{\alpha N + k - j} \binom{(1-\alpha_n)N}{(1-\alpha_n)N - k} \varepsilon^{\alpha N + k - j} (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{(1-\alpha)N-k}$$

The remaining transition probabilities, $p_{11}, p_{12}, p_{22}, p_{21}, p_{mm}$ and p_{m2} are given in the appendix.

The time spent in the neighborhood B_m can be obtained from the transition probabilities alone.

¹⁸Consider that a minimum of 20 mutations are required to leave B_m and enter B_1 . These 20 mutations must all be T_2 agents switching from their best response s_2 to s_1 . Then, as each T_1 mutation 'cancels out' a T_2 mutation, the difference in the number T_2 and T_1 mutations must be at least 20 for the process to jump from B_m to B_1 .

Lemma 3

$$\pi_m = \frac{1}{1 + \frac{\frac{p_{m2} (p_{12} + 1) + p_{21} p_{m1}}{p_{2m} p_{1m}}}{\frac{p_{12} + 1 + p_{21}}{p_{1m}}} + \frac{\frac{p_{m1} (p_{21} + 1) + p_{12} p_{m2}}{p_{1m} p_{2m}}}{\frac{p_{21} + 1 + p_{12}}{p_{2m}}}}$$

Proof. See the appendix. ■

We divide the state space into neighborhoods as it allows the opportunity to obtain results with vanishing and non-vanishing noise. To achieve such results we now define the two most important terms of the paper.

For analysis the dynamics under vanishing noise we have the familiar notion of stochastic stability.

Definition 1 An equilibrium E_i is defined as being *stochastically stable* if ¹⁹

$$\lim_{\varepsilon \rightarrow 0} \pi_i > 0$$

In order to investigate the dynamics under non-vanishing noise we introduce popular stability.

Definition 2 An equilibrium E_i is defined as being *popularly stable* at noise level ε if there exists ε such that

$$\lim_{N \rightarrow \infty} \pi_m > 0$$

As noise is not allowed to vanish for popular stability, the process is always dynamic and therefore no single state is selected as the population increases without limit. Instead an equilibrium's basin of attraction is selected as we increase the population size. Although individual noise does not vanish, on an aggregate level the noise of the population does reduce to zero due to the law of large numbers, thus popular stability can

¹⁹In fact, for vanishing noise all the time is spent at the single state $z = \alpha N$. I have used π_i here for a clearer comparison of the two limiting techniques. Also, the usual definition stochastic stability states $\lim_{\varepsilon \rightarrow 0} \pi_z > 0$, for simplicity i wish to focus only on when *all* the time is spent in one area.

select a unique neighbourhood as N increases without bound.²⁰

The main aim of this paper is to investigate whether analysis of stochastic stability consistently yields the same conclusions as popular stability when ε is *small*.²¹ In order to test this we focus on the conditions for which the long run location of the process is near E_m , for both limiting techniques. We begin with the simpler case of stochastic stability.

Proposition 1 *Under vanishing noise, E_m will be stochastically stable if and only if its basin B_m occupies at over half the markov space.*

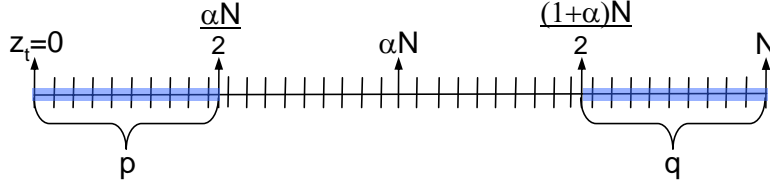
$$\lim_{\varepsilon \rightarrow 0} \pi_m > 0 \text{ iff } p \leq \frac{\alpha}{2} \text{ and } q \geq \frac{1+\alpha}{2}.$$

Analysis under vanishing noise can be seen as simply counting and comparing the number of mutations needed to escape each basin. And so in order for E_m to be stochastically stable it must take more mutations to escape B_m (to B_1 or B_2) than any other adjacent basin escape. Let the transition (jump) from basin i into basin j be denoted by B_{ij} . Then consider a scenario where the basin escape B_{m1} requires just one less mutation than B_{1m} , for some N . Then in the limit of $\varepsilon \rightarrow 0$, the 'cost' of this one extra mutation becomes infinitely large, overwhelming any other forces that could be in effect and ensuring that E_m is stochastically unstable. I shall refer to this force as the *basin size effect*. When obtaining results with vanishing noise the basin size effect is all that matters.

This can be seen by the illustration of Proposition 1, the shaded area representing the range of values that both p and q must take in order for E_m to be stochastically stable.

²⁰I would like to thank D.Myatt for this intuition.

²¹The question of what constitutes small noise has no simple answer. With further experimental data we could perhaps replace the word small with realistic.



And the basin size effect remains a very powerful force when we examine popular stability. To see this consider that the basin escape B_{m1}^{\rightarrow} requires 3 less mutations than B_{1m}^{\leftarrow} for some \tilde{N} . Then at twice this population size B_{m1}^{\rightarrow} now requires 6 less mutations than B_{1m}^{\leftarrow} . 24 less at $4\tilde{N}$, and so on. Therefore in a boundedly rational population, the magnitude of the basin size effect rises linearly with N . And so it would seem that basin sizes again will be all that determines equilibrium selection.

However, there are other forces at work which are overwhelmed under vanishing noise, but have the ability under non-vanishing noise to alter selection against the basin size effect.

The first I shall call *the combination effect*. At any state in B_2 all N agents could experiment with s_1 , while in B_m there are only αN agents able to experiment with s_2 . This contributes towards there being many more combinations of mutations available for a B_{2m}^{\rightarrow} jump than B_{m2}^{\leftarrow} . And as N rises this combinational difference also increases. In fact, this effect alone results in different popular and stochastic stability results. It is not hard to see that when a difference in equilibrium selection does occur it is popular stability that favors E_m .

The second effect I will call the *dis-coordination effect*. This is the effect, when in B_m only, of simultaneous mutations from both types canceling each other out, tending to make any B_m escape less probable at higher levels of noise. As N increases, the possible number of opposing mutations

to any jump from B_m also increases. Again this effect strengthens E_m under small non-vanishing noise, and has the potential to change selection away from the stochastically stable equilibrium.

However, when trying to obtain popular stability results two main problems appear. Firstly, when analysing with vanishing noise, one may consider only the basin jumps requiring the minimum number of mutations, as all other possible jumps become negligible in the limit of $\varepsilon \rightarrow 0$. But under non-vanishing noise in the limit of $N \rightarrow \infty$, the probability of other possible jumps do not become negligible and so need to be included in the analysis. For instance, when analyzing the probability of jumping from B_{1m} , one must consider the probability of jumping from B_1 to *any* state in B_m . As there exists many states in B_m , the number increasing in N , the calculation and analysis of basin escape probabilities can be complex. The second main problem is that the positive probability of simultaneous mutations from both types complicates the basin escape probabilities from B_m further. These are not straight binomial probabilities, but the net of two binomials.

Such complications make the derivation of precise critical values (p, q, α and ε) for particular equilibrium selection under positive noise a complex task. However using the following lemma and lemma 3 something can be said.

Lemma 4 *Let $Pr(S_n > r) = \sum_{v=0}^{\infty} b(r+v; v, l)$ where $b(r+v; v, l)$ is the binomial probability of exactly $r+v$ successes from n trials with l being the probability of a success. Then*

$$P(S_n \geq r) \leq b(r; n, l) \frac{r(1-l)}{r-nt} \forall l < r.$$

Proof. This is a standard result. For example see Feller p.151. ■

Lemma 4 allows us to outline the general conditions for E_m to be popu-

larly stable.

Proposition 2 $\lim_{N \rightarrow \infty} \pi_m \rightarrow 1$ if and only if

$$\begin{cases} \lim_{N \rightarrow \infty} \frac{p_{m2}}{p_{2m}} = 0 \text{ and} \\ \lim_{N \rightarrow \infty} \frac{p_{m1}}{p_{1m}} = 0 \forall \epsilon < \min(p, 1 - q). \end{cases}$$

Proof. See the appendix. ■

Proposition 2 essentially explains that as p_{12} and p_{21} are relatively negligible, if the inflows into B_m progressively dominate outflows as the population increases, then the process will spend all its time in B_m .

We can now do more than just look at stochastic stability as we are able to determine a condition for the co-existence equilibrium to be popularly stable.

Proposition 3 If

$$\begin{aligned} x(\alpha, p, \epsilon) &= \frac{\epsilon^{2p-\alpha}(1-\epsilon)^{1-2p}(\alpha-p)^{\alpha-p}}{\alpha^\alpha(1-p)^{(1-p)}} \geq 1 \text{ and} \\ y(\alpha, q, \epsilon) &= \frac{\epsilon^{2q-(1+\alpha)}(1-\epsilon)^{1-2q}(q-\alpha)^{q-\alpha}}{(1-\alpha)^{(1-\alpha)}q^q} \geq 1 \end{aligned}$$

then the time spent in B_m will be greater than that spent in B_1 or B_2 for $N > \tilde{N}$, p, q, α and $\epsilon > 0$.

For increasingly large N , if the two above conditions are satisfied then

$$\lim_{N \rightarrow \infty} \pi_m = 1$$

and so E_m is popularly stable for p, q, α and noise level ϵ .

Proof. See Appendix ■

It should be understood that x and y are not the exact critical points determining equilibrium selection at noise level ϵ , for some $x < 0$ and $y < 0$ it is still very possible that $\lim_{N \rightarrow \infty} \pi_m \rightarrow 1$. Analysis in Proposition 3 excludes the dis-coordination effect for tractability, and so does not capture the full strength of the co-existence equilibrium under small

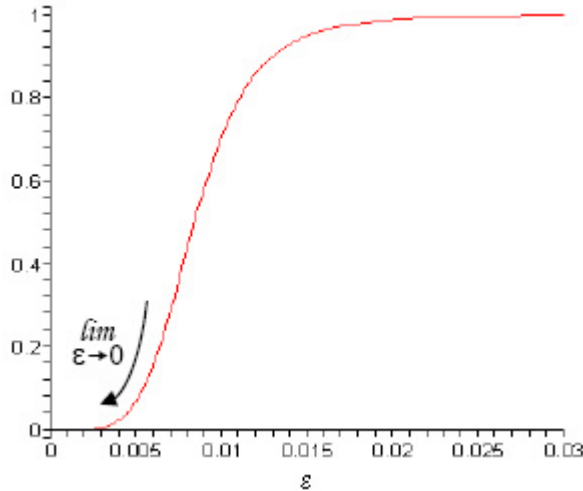
positive noise in an increasing population.

However Proposition 3 does allow us to establish our main result, that results of a process where noise vanishes completely can say very little of a population whose bounded rationality is innate.

Consider again the example in section 3 where $p = \frac{3}{11} = 0.2727$ and $q = \frac{8}{11} = 0.7272$.

T_1, T_2	s_1	s_2
s_1	8, 7	4, 0
s_2	0, 4	7, 8

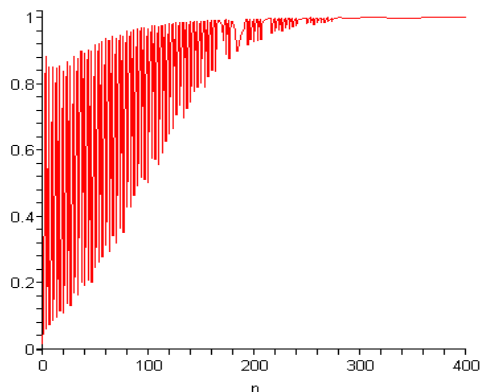
Here B_m contains less than half the markov space as both $p > \frac{\alpha}{2} = 0.25$ and $q < \frac{1+\alpha}{2} = 0.75$, from Proposition 2 we see that E_m is stochastically *unstable*. The calculations concur.



However, with a *small* non-vanishing experimentation rate of one in a 100 periods, $\varepsilon = 0.01$, we have it that

$$\begin{aligned}
 x(\alpha, p, \varepsilon) &= \frac{0.01^{2 \cdot (0.273) - 0.5} (1 - 0.01)^{1 - 2 \cdot (0.273)} (0.5 - 0.273)^{0.5 - 0.273}}{0.5^{0.5} (1 - 0.273)^{(1 - 0.273)}} \\
 &= y(\alpha, q, \varepsilon) = 1.0257 > 1
 \end{aligned}$$

and from Proposition we can see that E_m is popularly stable for $\varepsilon = 0.01$, and so the population will spend *all* its time in B_m as $N \rightarrow \infty$. The calculations agree.



Indeed, there exists a range of preferences for a large population where vanishing noise and small non-vanishing noise yield completely different results.

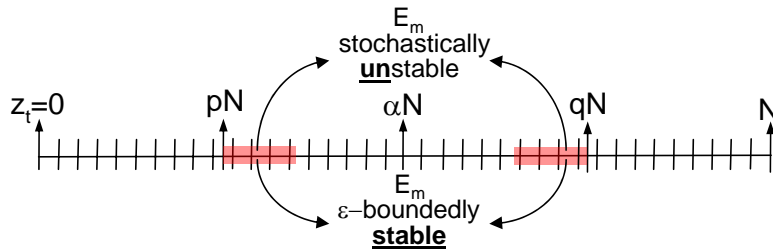
Theorem

For each $\varepsilon \in (0, \min(\frac{\alpha}{2}, \frac{1+\alpha}{2}))$ there exists a range of preferences corresponding to $p \in [\frac{\alpha}{2}, \frac{\alpha}{2} + \tau(\varepsilon))$ and $q \in [\frac{1+\alpha}{2}, \frac{1+\alpha}{2} - \tau(\varepsilon))$, $\tau(\varepsilon) > 0$, such that

$$\begin{cases} \lim_{\varepsilon \rightarrow 0} \pi_m = 0 & \forall \alpha, N, \text{ but} \\ \lim_{N \rightarrow \infty} \pi_m = 1 & \forall \alpha. \end{cases}$$

Proof. Please see appendix. ■

The theorem can be seen on the 1-dimensional diagram below, where the shaded region indicates for a given $\varepsilon > 0$ the range of p and q where the two limiting techniques give completely different results.



Corollary In the range $p \in [\frac{\alpha}{2}, \frac{\alpha}{2} + \tau(\varepsilon))$ and $q \in [\frac{1+\alpha}{2}, \frac{1+\alpha}{2} - \tau(\varepsilon))$, it is always popular stability that favors E_m as the long run equilibrium of the process, where as the stochastically stable state selects either E_1 or E_2 .

Proof. As $\tau(\varepsilon) > 0$, this follows straight from the theorem. ■

And so the technique of vanishing noise analysis does not show the true potential for boundedly rational populations to be caught in the neighborhood of the co-existence equilibria, often the worst place to be for the society to be. Therefore polymorphic states may play a much larger role than vanishing noise analysis would let us believe.

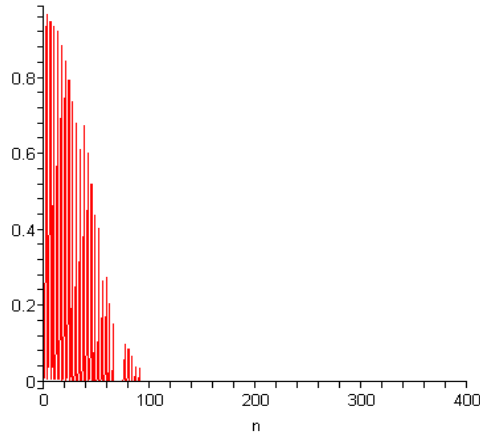
2.5 Calculations

In this section we examine a sample of calculations with $\rho = 0.5$. As will be seen, the more noisy a population becomes, the smaller B_m needs to be in order for the process to spend all its time there in a large population. Furthermore, we investigate a minority group who have strong preferences for their chosen strategy. In terms of social welfare, E_m is the worst equilibria for the society in every example.

Consider a population where there exists a less intense difference in T_1 and T_2 's preferences than in section 3 such that $\beta_1 = \gamma_2 = 1.58$ and $\beta_2 = \gamma_1 = 1$, giving $p = 0.286$ and $q = 0.714$.

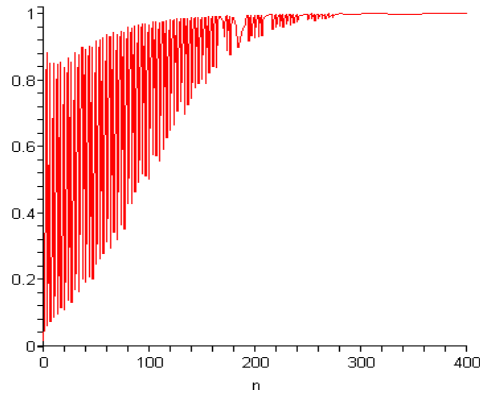
Now at the experimentation rate of one per hundred periods, $\varepsilon = 0.01$, $\alpha = 0.5$ and N increasing we see that for large N the process will in fact spend none of its time in the basin of E_m .

Note that $x(\frac{1}{2}, 0.286, 0.01) = y(\frac{1}{2}, 0.714, 0.01) = 0.93 < 1$.

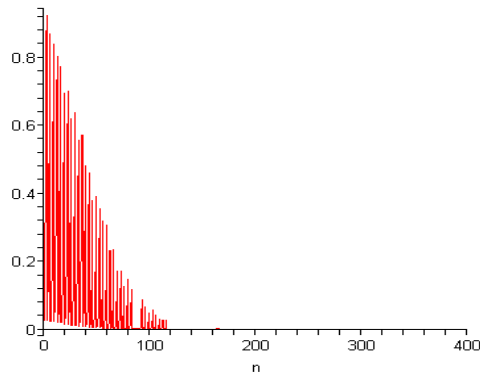


However, at a slightly larger mutation rate of 1 in 25, $\varepsilon = 0.04$, the population will again spend all of its time near E_m as the population grows large.

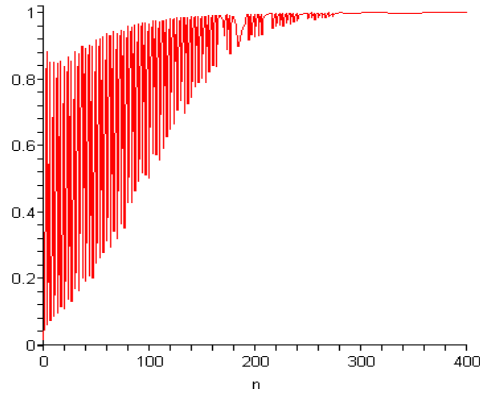
Here $x(\frac{1}{2}, 0.286, 0.01) = y(\frac{1}{2}, 0.714, 0.01) = 1.01 > 1$.



But as the difference of T1 and T2's preferences become less intense still with $\beta_1 = \gamma_2 = 1.53$ and $\beta_2 = \gamma_1 = 1$, with non-vanishing noise $\varepsilon = 0.04$, again the process spends almost none of its time in the basin of E_m for large N.



Indeed the trend continues as when noise increases to $\varepsilon = 0.07$, E_m once more becomes the long run location of the population.

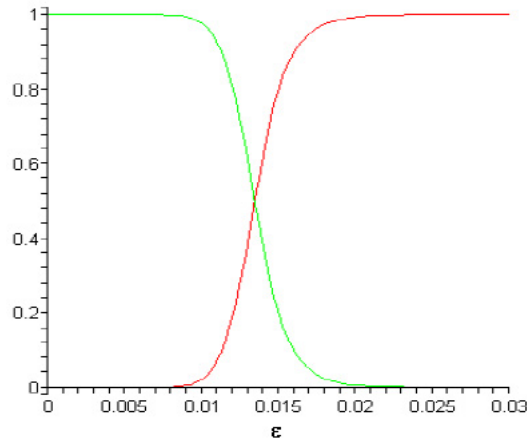


Consider that you are an observer, perhaps a member of a governing body who has no knowledge of the preferences of the population. Then one interpretation of the above trend is that the more noisy a population is, the more likely we are to find a large population spending nearly all of its time in a state of inefficient strategy co-existence.

2.5.1 Minority groups

So far our examples have focused upon an equal number of T_1 and T_2 agents with identically asymmetric preferences, but the model can easily lend itself to the analysis of minority groups. We again find that vanishing noise results can be a misleading portrayal of a boundedly rational population.

Consider twice as many T_2 agents than T_1 agents in a population of 90, but allow T_1 agents to have a stronger preference for their favoured strategy such that $\beta_1 = 1.83$, $\gamma_1 = 1$ and $\gamma_2 = 1.3$, $\beta_2 = 1$.



where π_m is given by the red line and π_2 by the green.

We see that results from vanishing noise analysis convey that the minority group's strong preferences have no influence over the state of the population, if they were indifferent between s_1 and s_2 they would be in the same position. However, in a population with some positive noise the minority group does have some sway in the population. At a small noise level of 0.02 we see that the population will spend almost all of its time near E_m , a significantly better location for the minority to be in.

2.6 Conclusion

This paper is motivated by the belief that conclusions drawn from vanishing noise results can often be surprisingly misleading when used to determine the nature of a truly boundedly rational population, even when this bounded rationality is small.

I investigate a typical KMRY type model with a population playing a 2x2 co-ordination game. The introduction of slight player heterogeneity creates a further steady state of co-existence of the two strategies.

By using a large population of agents who play a best response each period with probability $1-\varepsilon$, I have been able to obtain results from both vanishing *and* non-vanishing noise techniques, and can therefore compare the two.

I find that the two do not yield the same results. Indeed, there exists a range of population preferences where the two methods produce *completely different conclusions*. Vanishing noise analysis telling us that the population will spend all of its time co-ordinating on one strategy, while under small non-vanishing noise the population will in fact always be close to the co-existence steady state.

The reason for the startling difference between the methods is that the limiting procedure of vanishing noise is somewhat overpowering. There are important forces at work in the population dynamics which are simply overwhelmed and ignored when noise completely vanishes. However these forces can be of influence when noise is small, even when the population is large, and this is why we observe the disparity in the results of the two techniques.

Given that there exists such a discrepancy, this paper seeks to highlight a dangerous trend in the past literature to conduct vanishing noise analysis

alone, with little consideration of how significantly results would change with just a small amount of non-vanishing noise. As more and more studies emphasize the irrationality of their human subjects, perhaps the focus in the literature should be turning to truly noisy populations.

2.7 Appendix

Proof of Lemma 1:

At E_1 total social welfare (sw) is given by $\alpha N \beta_1 (N-1) + (1-\alpha) N \gamma_1 (N-1)$.

At E_m , $z = \alpha N$, $sw = \alpha N \beta_1 (\alpha N - 1) + (1-\alpha) N \gamma_1 ((1-\alpha)N - 1)$.

Therefore by setting $\alpha_1 = \gamma_2$ we have it that $E_m^{sw} > E_1^{sw}$ iff $2\gamma_2\alpha^2 + \gamma_2 - 3\gamma_2\alpha - \gamma_1 + \gamma_1\alpha > 0$.

a) At $\alpha = 0.5$, $E_m^{sw} - E_1^{sw} = -0.5 \gamma_1 < 0$ such that $E_1^{sw} > E_m^{sw}$ for all γ_1, β_2 where $\beta_1 \simeq \gamma_1$ and $\alpha = 0.5$.

Similarly at $\alpha = 0.5$, $E_m^{sw} - E_2^{sw} = -0.5 \beta_2 < 0$ such that $E_2^{sw} > E_m^{sw}$ for all γ_1, β_2 where $\beta_1 \simeq \gamma_1$ and $\alpha = 0.5$.

b) For any α , β_1 and γ_2 let $\gamma_1 = \beta_2 = 0$.

Then $E_m^{sw} > E_1^{sw}$ iff $\alpha < \frac{d}{a+d}$.

Similarly $E_m^{sw} > E_2^{sw}$ iff $\alpha > \frac{d}{a+d}$.

Therefore there will always exist another equilibrium which is superior to E_m in terms of total social welfare. ■

Proof of Lemma 3

The stationary equations of the 3 state markov process are given by

$$1. \pi_1 = \pi_1 p_{11} + \pi_2 p_{21} + \pi_m p_{m1}; \quad 2. \pi_2 = \pi_1 p_{12} + \pi_2 p_{22} + \pi_m p_{m2};$$

$$3. \pi_m = \pi_1 p_{1m} + \pi_2 p_{2m} + \pi_m p_{mm}; \quad 4. \pi_1 + \pi_2 + \pi_m = 1,$$

and also note,

$$5. p_{11} + p_{12} + p_{1m} = 1, \quad 6. p_{21} + p_{22} + p_{2m} = 1, \quad 7. p_{m1} + p_{m2} + p_{mm} = 1.$$

$$1 \Rightarrow \pi_1(1 - p_{11}) = \pi_2 p_{21} + \pi_m p_{m1}, \text{ and}$$

$$2 \Rightarrow \pi_2 = \frac{\pi_1 p_{12} + \pi_m p_{m2}}{1 - p_{22}}.$$

$$\text{Therefore 1 and 2} \Rightarrow \pi_1 = \left[\frac{\frac{p_{m1}(1-p_{22}) + p_{m2}}{p_{21}}}{(1-p_{11})(1-p_{22}) - p_{12}} \right], \quad \pi_2 = \left[\frac{\frac{p_{m1}(p_{21}+1) + p_{12} p_{m2}}{p_{1m} p_{2m}}}{\frac{p_{21}}{p_{2m}} + 1 + \frac{p_{12}}{p_{1m}}} \right] \pi_m,$$

from substituting in 5 and 6 and rearranging.

Symmetrically we have $\pi_2 = \left[\frac{\frac{p_{m2}}{p_{2m}} (\frac{p_{12}}{p_{1m}} + 1) + \frac{p_{21}}{p_{2m}} \frac{p_{m1}}{p_{1m}}}{\frac{p_{12}}{p_{1m}} + 1 + \frac{p_{21}}{p_{2m}}} \right] \pi_M$

From substituting both expressions into 4 we obtain our result. ■

Proof of Proposition 1

First consider $\frac{p_{m2}}{p_{2m}}$.

$$\begin{aligned}
p_{m2} &= \sum_{j=0}^{[pN+\delta]_-} \sum_{k=\max\{j-\alpha N, 0\}}^{\min\{j, N-i\}} \binom{\alpha N}{\alpha N - [pN + \delta]_-} \binom{N}{(1-\alpha)N - k} \\
&\quad \varepsilon^{\alpha N + k - j} (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{N-\alpha N - k} \\
&= \varepsilon^{\alpha N - [pN + \delta]_-} \binom{\alpha N}{\alpha N - [pN + \delta]_-} (1-\varepsilon)^{[pN + \delta]_-} (1-\varepsilon)^{(1-\alpha)N} + \\
&\quad \sum_{j=1}^{[pN+\delta]_-} \sum_{k=0}^j \binom{\alpha N}{\alpha N - [pN + \delta]_-} (1-\varepsilon)^{[pN + \delta]_-} \binom{N}{(1-\alpha)N - k} \varepsilon^{\alpha N + k - j} \\
&\quad (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{N-\alpha N - k} \\
&\quad + \sum_{k=1}^j \binom{\alpha N}{\alpha N - [pN + \delta]_-} \binom{N}{(1-\alpha)N - k} \varepsilon^{\alpha N + k - j} \\
&\quad (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{N-\alpha N - k} \\
&\equiv \varepsilon^{\alpha N - [pN + \delta]_-} [\rho f(1-\varepsilon) + \rho f(\varepsilon)]
\end{aligned}$$

where ρ is some function independent of ε , and,

$$\begin{aligned}
p_{2m} &= \sum_{i=[pN+\delta]_+}^{[qN+\zeta]_-} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \\
&= \varepsilon^{[pN+\delta]_+} [(N, [pN + \delta]_+)(1-\varepsilon)^{N-[pN+\delta]_+} + \sum_{i=[pN+\delta]_++1}^{[qN+\zeta]_-} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}] \\
&\equiv \varepsilon^{[pN+\delta]_+} [\rho f(1-\varepsilon) + \rho(\varepsilon)].
\end{aligned}$$

And so,

$$\frac{p_{m2}}{p_{2m}} = \frac{\varepsilon^{\alpha N - [pN + \delta]_-} [\rho f(1 - \varepsilon) + \rho f(\varepsilon)]}{\varepsilon^{[pN + \delta]_+} [\rho f(1 - \varepsilon) + \rho f(\varepsilon)]} = \varepsilon^{(\alpha - 2p)N - 2\delta + (\gamma_1 - \gamma_2)} \frac{[\rho f(1 - \varepsilon) + \rho f(\varepsilon)]}{[\rho f(1 - \varepsilon) + \rho f(\varepsilon)]}$$

by letting $[x]_- = x - \gamma_1$ and $[x]_+ = x + \gamma_2$ where $\gamma_1, \gamma_2 < 1$ for any x .

Therefore

$$\lim_{\varepsilon \rightarrow 0} \frac{p_{m2}}{p_{2m}} = \begin{cases} 0 & \text{if } p < \frac{\alpha}{2} - \frac{\delta + (\gamma_1 - \gamma_2)}{N} \\ \infty & \text{if } p > \frac{\alpha}{2} - \frac{\delta + (\gamma_1 - \gamma_2)}{N} \end{cases}$$

which essentially corresponds to

$$\lim_{\varepsilon \rightarrow 0} \frac{p_{m2}}{p_{2m}} = \begin{cases} 0 & \text{if } p < \frac{\alpha}{2} \\ \infty & \text{if } p > \frac{\alpha}{2}. \end{cases}$$

Similarly we have

$$\lim_{\varepsilon \rightarrow 0} \frac{p_{m1}}{p_{1m}} = \begin{cases} 0 & \text{if } q > \frac{1 + \alpha}{2} \\ \infty & \text{if } q < \frac{1 + \alpha}{2}. \end{cases}$$

Therefore as $\lim_{\varepsilon \rightarrow 0} \frac{p_{21}}{p_{2m}} = 0$ and $\lim_{\varepsilon \rightarrow 0} \frac{p_{12}}{p_{1m}} = 0 \forall N$, from lemma 3 we have it that $\lim_{\varepsilon \rightarrow 0} \pi_m = 1$ requires $p < \frac{\alpha}{2}$ and $q > \frac{1 + \alpha}{2}$

Proof of Proposition 2

Recall from lemma 1 that

$$\pi_M = \frac{1}{1 + \frac{\frac{p_{M1} (p_{21} + 1) + p_{12} p_{M2}}{p_{1M} p_{2M}} + \frac{p_{M2} (p_{12} + 1) + p_{21} p_{M1}}{p_{2M} p_{1M}}}{\frac{p_{21} + 1 + p_{12}}{p_{2M}} + \frac{p_{12}}{p_{1M}}}} \equiv \frac{1}{1 + \pi_1 + \pi_2}.$$

First note that $\frac{p_{21}}{p_{2M}} < c \forall \varepsilon < p$, $\varepsilon < 1 - q$, α and N , where c is some constant.

To see this first define

$$p_{2m}^j = \binom{N}{[pN+\delta]_++j} \varepsilon^{[pN+\delta]_++j} (1-\varepsilon)^{N-[pN+\delta]_++j} \text{ and}$$

$$p_{21}^j = \binom{N}{[qN+\zeta]_++j} \varepsilon^{[qN+\zeta]_++j} (1-\varepsilon)^{N-[qN+\zeta]_++j}$$

and then consider

$$\frac{p_{21}}{p_{2m}} = \frac{p_{21}^0 + p_{21}^1 + \dots + p_{21}^{N-[qn+\zeta]_+}}{p_{2m}^0 + p_{2m}^1 + \dots + p_{2m}^{N-[pn+\delta]_+}} < \frac{p_{21}^0 [qN+\zeta]_+ (1-\varepsilon)}{p_{2m}^0 [qN+\zeta]_+ - N\varepsilon}$$

using lemma 1.

$$\text{As } \lim_{N \rightarrow \infty} \frac{[qN+\zeta]_+ (1-\varepsilon)}{[qN+\zeta]_+ - N\varepsilon} = \frac{q(1-\varepsilon)}{q-\varepsilon} < c_1 \text{ and } \frac{p_{21}^0}{p_{2m}^0} < 1 \quad \forall \varepsilon < p,$$

we have it that $\frac{p_{21}}{p_{2M}} < c \quad \forall \varepsilon < p$. And symmetrically $\frac{p_{21}}{p_{2M}} < \acute{c} \quad \forall \varepsilon < p$.

Therefore if $\frac{p_{m2}}{p_{2m}} \rightarrow 0$ and $\frac{p_{m1}}{p_{1m}} \rightarrow 0$ as $N \rightarrow \infty$, then

$$\lim_{N \rightarrow \infty} \pi_1 = \lim_{N \rightarrow \infty} \frac{\frac{p_{m1}}{p_{1m}} \left(\frac{p_{21}}{p_{2m}} + 1 \right) + \frac{p_{12}}{p_{1m}} \frac{p_{m2}}{p_{2m}}}{\frac{p_{21}}{p_{2m}} + 1 + \frac{p_{12}}{p_{1m}}} = \frac{0+0c+0\acute{c}}{0+1+0} = 0.$$

Similarly $\lim_{N \rightarrow \infty} \pi_2 = 0$.

And therefore $\lim_{N \rightarrow \infty} \pi_m = \frac{1}{1+0+0} = 1$. ■

Proof of Proposition 3

First consider when $\frac{p_{2m}}{p_{m2}}$ is rising with N.

By considering that the probability of a basin escape is at all times higher under a constraint that no opposing T_2 mutations can occur in a period, we can deduce an upper bound

$$p_{m2} = \sum_{j=0}^{[pN+\delta]_-} \sum_{k=\max\{j-\alpha N, 0\}}^{\min\{j, N-i\}} \binom{\alpha N}{\alpha N - [pN + \delta]_-} ((1-\alpha)N - k, N) \varepsilon^{\alpha N + k - j} (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{N-\alpha N - k}$$

$$< \sum_{j=0}^{[pN+\delta]_-} \binom{\alpha N}{\alpha N - [pN + \delta]_- + j} \varepsilon^{\alpha N - [pN + \delta]_- + j} (1-\varepsilon)^{[pN+\delta]_- - j}$$

$$< \binom{\alpha N}{\alpha N - [pN + \delta]_-} \varepsilon^{\alpha N - [pN + \delta]_-} (1-\varepsilon)^{[pN+\delta]_-} \frac{[pN + \delta]_- (1-\varepsilon)}{[pN + \delta]_- - N\varepsilon}$$

$\forall \varepsilon, p, \alpha$ and N ,

the third part coming from lemma 4.

Also,

$$\begin{aligned} p_{2m} &= \sum_{j=0}^{[pN+\delta]_-} \binom{N}{[pN+\delta]_+ + j} \epsilon^{[pN+\delta]_- + j} (1-\epsilon)^{N-[pN+\delta]_- - j} \\ &> \binom{N}{[pN+\delta]_+} \epsilon^{\alpha N - [pN+\delta]_-} (1-\epsilon)^{[pN+\delta]_-} \forall \epsilon, p, \alpha \text{ and } N. \end{aligned}$$

And so,

$$\begin{aligned} \frac{p_{2m}}{p_{m2}} &> \frac{\binom{N}{[pN+\delta]_+} \epsilon^{[pN+\delta]_-} (1-\epsilon)^{N-[pN+\delta]_-}}{\binom{\alpha N}{\alpha N - [pN+\delta]_-} \epsilon^{\alpha N - [pN+\delta]_-} (1-\epsilon)^{[pN+\delta]_-} \frac{[pN+\delta]_- (1-\epsilon)}{[pN+\delta]_- - N\epsilon}} \\ &\forall \epsilon, p, \alpha \text{ and } N. \end{aligned}$$

And so when the right-hand side of this inequality, label it λ , increases without bound as $N \rightarrow \infty$, then so must $\frac{p_{2m}}{p_{m2}} \rightarrow \infty$ as $N \rightarrow \infty$. And so we now investigate under which parameter values the right hand side increases without bound as N rises.

Now,

$$\begin{aligned} \lambda &= \frac{\binom{N}{[pN+\delta]_+} \epsilon^{[pN+\delta]_+} (1-\epsilon)^{N-[pN+\delta]_+}}{\binom{\alpha N}{\alpha N - [pN+\delta]_-} \epsilon^{\alpha N - [pN+\delta]_-} (1-\epsilon)^{[pN+\delta]_-} \frac{[pN+\delta]_- (1-\epsilon)}{[pN+\delta]_- - N\epsilon}} \\ &= \epsilon^{[pN+\delta]_- + [pN+\delta]_+ - \alpha N} (1-\epsilon)^{N-[pN+\delta]_+ - [pN+\delta]_-} \\ &\quad \frac{N! (\alpha N - [pN+\delta]_-)! [pN+\delta]_-!}{(\alpha N)! (N - [pN+\delta]_+)! [pN+\delta]_+!} \frac{[pN+\delta]_- - N\epsilon}{[pN+\delta]_- (1-\epsilon)} \\ &= \epsilon^{[pN+\delta]_- + [pN+\delta]_+ - \alpha N} (1-\epsilon)^{N-[pN+\delta]_+ - [pN+\delta]_-} \\ &\quad \frac{N! (\alpha N - [pN+\delta]_-)!}{(\alpha N)! (N - pN + \delta]_+)! ([pN+\delta]_+ + 1)} \frac{[pN+\delta]_- - N\epsilon}{[pN+\delta]_- (1-\epsilon)} \end{aligned}$$

$$\text{as } \frac{[x]_-!}{[x]_+!} = \frac{1}{[x]_+ + 1}.$$

And

$$\begin{aligned} \lambda &= \epsilon^{[pN+\delta]_- + [pN+\delta]_+ - \alpha N} (1-\epsilon)^{N-[pN+\delta]_+ - [pN+\delta]_-} \\ &\quad \frac{N! (\alpha N - [pN+\delta]_-)! (N - [pN+\delta]_-)}{(\alpha N)! (N - pN + \delta]_-)! ([pN+\delta]_+ + 1)} \frac{[pN+\delta]_- - N\epsilon}{[pN+\delta]_- (1-\epsilon)} \end{aligned}$$

as $(x - [k]_+)! = (x - [k]_- - 1)! = \frac{(x - [k]_-)!}{x - [k]_-}$.

Using $[x]_- = x - \gamma_1$ and $[x]_+ = x + \gamma_2$ where $\gamma_1, \gamma_2 < 1$ for any x , and taking natural logarithms of both sides we obtain

$$\begin{aligned} \ln \lambda &= ((2p - \alpha)N - 2\delta + \gamma_1 - \gamma_2) \ln \varepsilon + ((1 - 2p)N - 2\delta + \gamma_1 - \gamma_2) \ln(1 - \varepsilon) \\ &\quad + \ln N! - \ln(\alpha N)! + \ln(N - pN + \delta - \gamma_1)! + \ln \frac{(1 - p)N + \delta - \gamma_1}{pN + \delta + \gamma_2 + 1} \\ &\quad + \ln(\alpha N - pN + \delta - \gamma_1)! + \ln \frac{(p - \varepsilon)N + \delta - \gamma_1}{(pN + \delta - \gamma_1)(1 - \varepsilon)}. \end{aligned}$$

As we shall be taking the limit of $N \rightarrow \infty$, we are able to make use of stirling's formula which states

$$\lim_{x \rightarrow \infty} \frac{\ln x!}{x \ln x - x} = 1$$

Substituting this in gives

$$\begin{aligned} \ln \lambda &= ((2p - \alpha)N - 2\delta + \gamma_1 - \gamma_2) \ln \varepsilon + ((1 - 2p)N - 2\delta + \gamma_1 - \gamma_2) \ln(1 - \varepsilon) \\ &\quad + N(\ln N - 1) + (\alpha N - pN + \delta - \gamma_1)(\ln(\alpha N - pN + \delta - \gamma_1) - 1) + \\ &\quad \ln \frac{N - pN + \delta - \gamma_1}{pN + \delta + \gamma_2 + 1} - \alpha N \ln(\alpha N - 1) - \\ &\quad (N - pN + \delta - \gamma_1) \ln((N - pN + \delta - \gamma_1) - 1) + \ln \frac{(p - \varepsilon)N + \delta - \gamma_1}{(pN + \delta - \gamma_1)(1 - \varepsilon)}. \end{aligned}$$

Now implementing

$$\ln(x + \xi) = \ln x + \frac{\xi}{x} - \frac{\xi^2}{2x^2} + \frac{\xi^3}{3x^3} - \dots$$

gives

$$\begin{aligned} \lambda &= (2p - \alpha)N \ln \varepsilon + (1 - 2p)N \ln(1 - \varepsilon) - \alpha N \ln \alpha - (1 - p)N \ln(1 - p) \\ &\quad + (\alpha - p)N \ln(\alpha - p) + N \ln N + (\alpha - p)N \ln N - \alpha N \ln N - (1 - p)N \ln N \\ &\quad - N - (1 - p)N + (\alpha - p)N + (\gamma_1 - 2\delta - \gamma_2) \ln \varepsilon + (\gamma_1 - 2\delta - \gamma_2) \ln(1 - \varepsilon) \\ &\quad - (1 - p)N \left(\frac{\delta - \gamma_1}{N} + \frac{(\delta - \gamma_1)^2}{N^2} + \dots \right) + (\alpha - p)N \left(\frac{\delta - \gamma_1}{N} - \frac{(\delta - \gamma_1)^2}{N^2} + \dots \right) \\ &\quad + \ln \frac{(1 - p)N + \delta - \gamma_1}{pN + \delta + \gamma_2 + 1} + \ln \frac{(p - \varepsilon)N + \delta - \gamma_1}{(pN + \delta - \gamma_1)(1 - \varepsilon)} + \gamma_1 \end{aligned}$$

or,

$$\begin{aligned} \lambda &= \ln\left[\frac{\varepsilon^{2p-\alpha}(1-\varepsilon)^{1-2p}(1-p)^{(1-p)}}{(\alpha-p)^{\alpha-p}\alpha^\alpha}\right]N + (\gamma_1 - 2\delta - \gamma_2) \ln \varepsilon + (\gamma_1 - 2\delta - \gamma_2) \ln(1-\varepsilon) \\ &\quad - (1-p)(\delta - \gamma_1 - \frac{(\delta - \gamma_1)^2}{N} + \dots) + (\alpha-p)(\delta - \gamma_1 - \frac{(\delta - \gamma_1)^2}{N} + \dots) \\ &\quad + \ln \frac{(1-p)N + \delta - \gamma_1}{pN + \delta + \gamma_2 + 1} + \ln \frac{(p-\varepsilon)N + \delta - \gamma_1}{(pN + \delta - \gamma_1)(1-\varepsilon)} + \gamma_1 \end{aligned}$$

In the limit as $N \rightarrow \infty$ we have it that

$$\ln \frac{(1-p)N + \delta - \gamma_1}{pN + \delta + \gamma_2 + 1} + \ln \frac{(p-\varepsilon)N + \delta - \gamma_1}{(pN + \delta - \gamma_1)(1-\varepsilon)} \rightarrow \ln \frac{(1-p)}{p} + \ln \frac{(p-\varepsilon)}{p(1-\varepsilon)} < c_2.$$

And so $\lim_{N \rightarrow \infty} \lambda \rightarrow \infty$, and therefore $\lim_{N \rightarrow \infty} \frac{p_{2m}}{p_{m^2}} \rightarrow \infty$, if and only if

$$\ln\left[\frac{\varepsilon^{2p-\alpha}(1-\varepsilon)^{1-2p}(\alpha-p)^{\alpha-p}}{\alpha^\alpha(1-p)^{(1-p)}}\right] > 0.$$

Symmetrical analysis yields that if

$$\begin{aligned} \ln\left[\frac{\varepsilon^{2q-(1+\alpha)}(1-\varepsilon)^{1-2q}(q-\alpha)^{q-\alpha}}{(1-\alpha)^{(1-\alpha)}q^q}\right] > 0 \text{ then,} \\ \lim_{N \rightarrow \infty} \frac{p_{1m}}{p_{m^1}} \rightarrow \infty. \end{aligned}$$

And so from Proposition 2, if both the above conditions are satisfied then

$$\lim_{N \rightarrow \infty} \pi_m \rightarrow 1. \blacksquare$$

Proof of Theorem

To prove the theorem first consider the following lemma.

Lemma 5 At $p = \frac{\alpha}{2}$ and $q = \frac{1+\alpha}{2}$, $\lim_{N \rightarrow \infty} \pi_m = 1 \forall \alpha$ and $0 < \varepsilon \leq \min(\frac{\alpha}{2}, \frac{1+\alpha}{2})$.

Proof.

First let us consider $x(\alpha, p, \varepsilon)$ from Proposition 3.

$$\begin{aligned} \text{At } p = \frac{\alpha}{2}, \text{ we have it that } x(\alpha, \frac{\alpha}{2}, \varepsilon) &= \frac{\alpha}{2} \ln \frac{\alpha}{2} - \alpha \ln \alpha - (1 - \frac{\alpha}{2}) \ln(1 - \frac{\alpha}{2}) + \\ &\quad (1 - \alpha) \ln(1 - \varepsilon) \\ &= -\frac{\alpha}{2} \ln 2\alpha - (1 - \frac{\alpha}{2}) \ln(1 - \frac{\alpha}{2}) + (1 - \alpha) \ln(1 - \varepsilon). \end{aligned}$$

Let $\varepsilon = \frac{\alpha}{2}$ for each α , then we have

$$x(\alpha, \frac{\alpha}{2}, \frac{\alpha}{2}) = -\frac{\alpha}{2} \ln 2\alpha - \frac{\alpha}{2} \ln(1 - \frac{\alpha}{2}) = -\frac{\alpha}{2} \ln(2\alpha - \alpha^2) > 0 \quad \forall \alpha \in (0, 1).$$

And as $\frac{\partial x(\alpha, \frac{\alpha}{2}, \varepsilon)}{\partial \varepsilon} < 0$ we have it that $x(\alpha, \frac{\alpha}{2}, \varepsilon) > 0 \quad \forall \varepsilon \leq \frac{\alpha}{2}$.

Symmetrically, at $q = \frac{1+\alpha}{2}$, $y(\alpha, \frac{1+\alpha}{2}, \varepsilon) > 0$ for any $\varepsilon < \frac{1+\alpha}{2}$.

Therefore $x(\alpha, \frac{\alpha}{2}, \varepsilon) > 0$ and $y(\alpha, \frac{1+\alpha}{2}, \varepsilon) > 0 \quad \forall \varepsilon < \min(\frac{\alpha}{2}, \frac{1+\alpha}{2})$, and by Proposition 3 we are done.

Now we can prove the theorem.

Proposition 2 shows that $\lim_{\epsilon \rightarrow 0} \pi_m = 0 \quad \forall \alpha, N$ for any $p > \frac{\alpha}{2}$ and/or $q < \frac{1+\alpha}{2}$.

Now consider $\lim_{N \rightarrow \infty} \pi_m = 1$ and $x(\alpha, p, \epsilon)$.

Let $\alpha = \frac{2p}{k}$ and consider $k \in [1, 2)$ such that $\alpha \in (p, 2p]$.

Fix p and consider α varying.

For each p fix ε at some $\bar{\varepsilon} \in (0, \min(p, \frac{1-2p}{2}))$.²²

Then,

$$\begin{aligned} x(\frac{2p}{k}, p, \bar{\varepsilon}) &= 2p(1 - k^{-1}) \ln \bar{\varepsilon} + (1 - 2p) \ln(1 - \bar{\varepsilon}) + (\frac{2p}{k} - p) \ln(\frac{2p}{k} - p) - \\ &\frac{2p}{k} \ln \frac{2p}{k} - (1 - p) \ln(1 - p). \text{ and so,} \\ \frac{\partial x(\frac{2p}{k}, p, \bar{\varepsilon})}{\partial k} &= \frac{2p}{k^2} \ln \bar{\varepsilon} - \frac{2p}{k^2} \ln(\frac{2p}{k} - p) - \frac{2p}{k^2} + \frac{2p}{k^2} \ln \frac{2p}{k} - \frac{2p}{k^2} = \frac{2p}{k^2} [\ln \bar{\varepsilon} - \ln(\frac{2p}{k} - \\ &p) + \ln \frac{2p}{k}] \end{aligned}$$

which is continuous $\forall k \in [1, 2)$ provided $\varepsilon > 0$.

At $k = 1$, $\frac{\partial x(\frac{2p}{k}, p, \bar{\varepsilon})}{\partial k} = 2p[\ln 2\bar{\varepsilon}] > -\infty \quad \forall \bar{\varepsilon} > 0$.

From lemma 5 we have $x(\frac{2p}{k}, p, \bar{\varepsilon}) > 0$ at $k = 1$, therefore for each ε there must exist some range of $k > 1$, $\tau(\varepsilon)$, where x is positive. I.E, there exists some range of $p \in [\frac{\alpha}{2}, \frac{\alpha}{2} + \tau(\varepsilon)]$ such that $x > 0$.

Symmetrical analysis gives $y > 0$ for some range of at least $q \in [\frac{1+\alpha}{2} - \tau(\varepsilon), \frac{1+\alpha}{2}]$ and from Proposition 3 we are done. ■

²²This comes from $\varepsilon \leq \min(\min(2p, \frac{1-2p}{2}), \min(p, \frac{1-p}{2}))$

The other markov probabilities are given by:

$$\begin{aligned}
p_{11} &= \sum_{i=0}^{N-[qN+\zeta]_+} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}. \\
p_{12} &= \sum_{i=N-[pN+\delta]_-}^N \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \\
p_{22} &= \sum_{i=0}^{[pN+\delta]_-} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}. \\
p_{2m} &= \sum_{i=[pN+\delta]_+}^{[qN+\zeta]_-} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}. \\
p_{mm} &= \sum_{j=[pN+\delta]_+}^{j=[qN+\zeta]_-} \sum_{k=\max\{j-\alpha N, 0\}}^{\min\{j, N-i\}} \binom{\alpha N}{\alpha N+k-j} ((1-\alpha)N-k) \\
&\quad \epsilon^{\alpha N+k-j} (1-\epsilon)^{j-k} \epsilon^k (1-\epsilon)^{N-\alpha N-k}. \\
p_{m2} &= \sum_{j=\alpha N-[pN+\delta]_-}^{\alpha N} \sum_{k=\max\{j-\alpha N, 0\}}^{\min\{j, N-i\}} \binom{\alpha N}{\alpha N+k-j} ((1-\alpha)N-k) \\
&\quad \epsilon^{\alpha N+k-j} (1-\epsilon)^{j-k} \epsilon^k (1-\epsilon)^{N-\alpha N-k}.
\end{aligned}$$

References Anwar, A. (1999) "On the Coexistence of Conventions (2002)".

Journal of Economic Theory.

Alan Beggs, (2002). "Large Deviations and Equilibrium Selection in Large Populations". Economics Series Working Papers 129, University of Oxford, Department of Economics.

Blume, L. (1999): "How Noise Matters" Working Paper.

Bergin, J., and B. Lipman (1996): "Evolution with State-Dependent Mutations", *Econometrica*, 64, 943-956.

Ellison, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047-71.

——— (2000): "Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution" *Review of Economic Studies*, 67,

17-45.

Granovetter, Mark. (1978). "Threshold Models of Collective Behavior." *American Journal of Sociology*, 83: 1420-1443.

Hojman D, (2004) "Interaction Structure and the Evolution of Conventions."

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

Kolstad, I. (2003): "Viscosity and dispersion in an evolutionary model of learning."

Myatt, D.P and C. Wallace (1998): "Adaptive Play by Idiosyncratic Agents," *Discussion Paper, Nuffield College, Oxford*.

Norman, T. (2003a): "The Evolution of Co-ordination under Inertia."

Robson, Arthur J. and Vega-Redondo, Fernando, (1996): "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *Journal of Economic Theory*, Elsevier, vol. 70(1), pages 65-92.

Samuelson, l., (1994):"Stochastic Stability in Games with Alternative Best Replies," *Journal of Economic Theory* 64 , 35-65.

Sandholm, W. H. (2006). "Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium". Unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2007). "Population games and evolutionary dynamics". Unpublished manuscript, University of Wisconsin.

Schelling, Thomas C. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1: 143-186.

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

(1990) "Stochastic Evolutionary Game Dynamics" with Dean P. Foster.

Theoretical Population Biology, 38, 219-232.

3 Three's a Crowd

Abstract

This paper presents a co-ordination game played by a heterogenous population containing several different types of agent. As such there are many equilibria where two strategies co-exist together in varying proportions. Firstly we are able to deduce the characteristics of the equilibria that the population will be located in under stochastic stability. Secondly we examine the entrance of a neutral group into a population consisting of two opposing groups and determine the equilibrium selection results. We find the introduction of a small neutral group can produce a welfare improving switch in equilibrium selection away from co-existence.

3.1 Introduction

Since the introduction of game theory itself a stubborn problem has existed, when a game produces more than one nash equilibria which one should be selected? A population's long run location will depend only on their prior beliefs, and can easily become stuck in an inefficient equilibrium unable to escape. The ground-breaking papers of Kandori, Mailath and Rob(1993, KMR) and Young(1993) made a great leap in addressing this problem. In a homogenous population playing a binary action co-ordination game KMRY introduce boundedly rational agents who occasionally experiment away from their best choice, giving the population movement and allowing the possibility of escaping equilibria. Then by allowing the experimentation rate to fall to 0 KMRY are able to select which of their two stable equilibria will be selected.

This paper seeks to expand on KMRY's work by introducing a general level of heterogeneity into KMRY's initial model, attempting to continue to link their work with the general binary-action games of the kind introduced by Schelling. KMRY consider a homogeneous population with two stable equilibria, in which all agents choose the same action. In this paper we investigate a population consisting of many different types of player who vary in their preferences for two strategies in a co-ordination game. For example one type of player may strongly favour a certain strategy while another type prefer the opposite strategy, there may be also be fairly indifferent players and so on within a single population. This increased level of heterogeneity creates many equilibria, two of which correspond to perfect co-ordination with all agents agreeing on one strategy as in KMRY. However the remainder of the equilibria are states of co-existence, where both strategies survive in the long run. These equilibrium states represent

a population in conflict as there are two separate groups content playing different strategies, which can often be to the detriment of society as a whole.

There are several possible applications for the model. For instance consider the competition between the apple macintosh and windows operating systems. With a large range of different preferences existing in the population of computer users, the long run outcome has been a co-existence of both systems. Another application could be a key vote in a society between two opposing issues, such as the U.S presidential election. Here voters prefer others to agree with them and there exists a large range of different views within the population. Members of the population can see voting ratios on a regular basis via polls and are likely to occasional change their mind with similarities to our model.

This paper is not the first to look at introducing some heterogeneity into players' preferences in KMRY type models. For instance Hehenkamp(2001) considers two populations with asymmetric preferences. Kolstadt(2003) also looks at a model with two separate populations with conflicting preferences. Hahn(1997) looks to apply KMRYs model to the battle of the sexes game without finding a robust equilibrium selection process. Myatt and Wallace (1998) have a heterogeneity in players' preferences as they consider players with idiosyncratic payoffs in the form of Harsanyian type payoff trembles. There are also several papers whose models possess co-existence equilibria such as Ahmed(1999) and Norman(2003). To the best of the author's knowledge the current paper is the first to look to general heterogeneity in player types within one KMRY type population.

The paper begins by introducing the model in section 2 in which we have several different player types within a population, each with their own

preference for two strategies of a co-ordination game, creating a many equilibria. The next section addresses which of the many equilibria the population will spend most of its time in the long run under stochastic stability. As such by allowing the experimentation rate to fall to 0 in a fixed population, for any general level of heterogeneity, we are able to determine the precise characteristics of the long term location of the population.

In section 4 we apply the results from the previous section to a specific case of a three type population. In particular we wish to consider a possibly divided society which initially consisted of just two types of players with opposing preferences, and then at a later date experienced a group of fairly neutral players enter the population. Firstly we determine the exact characteristic equations of the stable equilibria for this population. Secondly, we show that in some cases only a small amount of neutral agents are required to upset a co-existence equilibria and push society to full agreement. We find when the introduction of neutral agents has this effect it is welfare improving for society as a whole. Section 5 concludes.

3.2 The Model

Consider a single population of N players consisting of m different types of players, type i being denoted by T_i , $i \in \{1, \dots, m\}$. The players are in a co-ordination game as in any period the more players in the population playing an agent's current strategy, the higher the agent's payoff. However the different types of players within the population receive different payoffs each period, due to their personal preferences. Each player has two possible strategies to choose from.

The payoff in any period t for a T_i agent playing strategy s_k , $\lambda_i^{s_k}$, is given by

$$\begin{aligned}\lambda_i^{s_1} &= \beta_i(z(t) - 1)^\rho \\ \lambda_i^{s_2} &= \gamma_i(N - z(t) - 1)^\rho\end{aligned}$$

where $z(t)$ represents the number of agents playing s_1 in period t and where $\rho \in \mathbb{R}^+$.²³

We have it that $\beta_i > \beta_j \forall i < j$. As such T_1 players have the strongest preference to co-ordinate on s_1 , followed by T_2 players, and so on. Indeed, T_1 players may require only a small proportion of the population to play s_1 in order for s_1 to be their best response. Similarly, $\gamma_i < \gamma_j \forall i < j$ indicates T_m players have the strongest preference to co-ordinate on s_2 .²⁴

²³For most applications we would have $\rho \in (0, 1]$ but we leave $\rho > 1$ open for generality.

²⁴We leave $\beta_1 < \gamma_1$ open as a possibility for generality.

3.2.1 The Dynamics of a Heterogenous Population

Our population consists of myopic players who consider the state of play will remain the same as the previous period, $z(t - 1)$. Therefore the best response of any T_i agent in any period t is²⁵

$$\begin{cases} s_1 & \text{if } z(t - 1) > \frac{1}{1 + (\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}}} N + \frac{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} - 1}{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} + 1} \equiv p_i N \\ s_2 & \text{Otherwise} \end{cases}$$

We shall continue by denoting N_i as the number of T_i agents in a population, such that $N = \sum_{i=1}^{i=m} N_i$. We define the proportion of T_i agents in the population as $\alpha_i = N_i/N$.²⁶ As agents consider only the aggregate amount of players choosing a strategy, $z(t) = \{0, 1, \dots, N\}$ defines the state of the process at any time t .

Two monomorphic stable points for the process exist, E_m where all agents choose to play $s_1(z = N)$ and E_0 where all choose $s_2(z = 0)$.²⁷

Furthermore, given

$$p_k < \sum_{i=1}^{i=k} \alpha_i < p_{k+1}$$

there exists $m - 1$ stable points of co-existence denoted by E_k ,

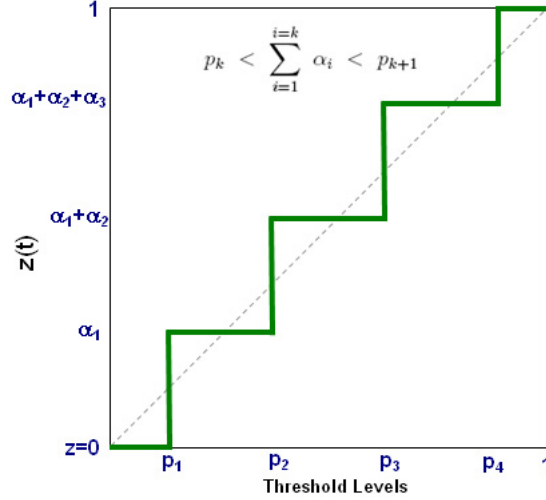
$k \in \{1, \dots, m - 1\}$ where a proportion $\sum_{i=1}^{i=k} \alpha_i$ agents will play s_1 and $\sum_{i=k+1}^{i=m} \alpha_i$ agents play s_2 . For the remainder of this paper we shall only consider cases where this constraint holds for all values of k .

The constraint is represented by the diagram below which shows the proportion of a four type population who will play s_1 given a proportion p_j played s_1 last period.

²⁵Such that $p_i = \frac{1}{1 + (\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}}} + \frac{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} - 1}{((\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} + 1)N}$.

²⁶We define $N_0 = 0$ and $\alpha_0 = 0$ for later analysis.

²⁷A notation change from the author's previous 'Noise Matters in Heterogenous Populations' chapter.



Overall there exists $m + 1$ possible stable states of the process, two of which are monomorphic, and $m - 1$ which are states of co-existence.

Let us continue by addressing the basins of attraction of each stable point and let the basin of attraction of E_k be denoted by B_k .

B_k is then defined as any state $z(t)$ in²⁸

$$\begin{cases} \{0, \dots, [p_1 N]_-\} & \text{for } k = 0 \\ \{[p_m N]_+, \dots, N\} & \text{for } k = m \\ \{[p_k N]_+, \dots, [p_{k+1} N]_-\} & \text{for } k \in \{1, \dots, m-1\} \end{cases}$$

For a population of agents who always play a myopic best response the long run location of the process depends upon the initial set up alone. However, as in KMRY, any agent can select a strategy other than its best response with probability ε each period, I shall describe such an event as a mutation.²⁹ As it is possible to jump from any given state to any

²⁸Defining $[x]_-$ as the as the nearest integer below or equal to x and $[x]_+$ as the nearest integer above or equal to x . If the proportion of s_1 played after strategies are chosen is within a certain basin region the process will remain in that basin for the beginning of the next period.

²⁹This system is often described as simultaneous revisions in the literature.

other state in one period, including itself, the process is irreducible and aperiodic, and therefore the markov chain is ergodic.

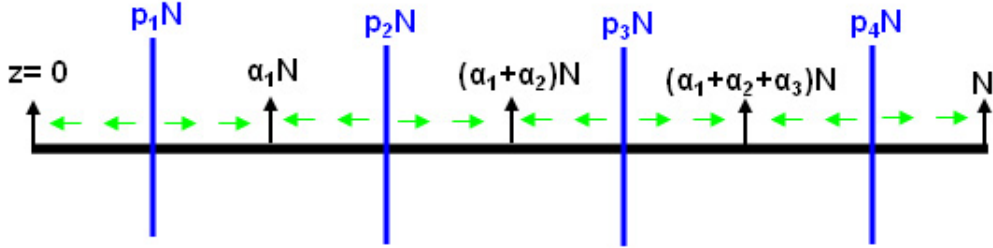
The probability of escaping B_0 to its immediate neighbouring basin B_1 in any period is given by the binomial probability

$$p_{01} = \sum_{i=[p_1 N]_+}^{[p_2 N]_-} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i}$$

Escaping from polymorphic states is a more complicated affair as simultaneous mutations from different types effectively cancel each other out, as such the probability of escaping B_k and entering B_{k-1} is given by ³⁰

$$p_{k,k-1} = \sum_{j=0}^{[p_k N]_-} \sum_{k=\max\{j-\sum_{i=0}^{i=k} \alpha_k N, 0\}}^{\min\{j, N-j\}} \binom{\sum_{i=0}^{i=k} \alpha_k N}{\sum_{i=0}^{i=k} \alpha_k N - [p_k N]_-} \binom{(1 - \sum_{i=0}^{i=k} \alpha_k) N}{(1 - \sum_{i=0}^{i=k} \alpha_k) N - k} \varepsilon^{\sum_{i=0}^{i=k} \alpha_k N + k - j} (1 - \varepsilon)^{j-k} \varepsilon^k (1 - \varepsilon)^{(1 - \sum_{i=0}^{i=k} \alpha_k) N - k}$$

The state space $z(t) = \{0, \dots, N\}$ is represented below, using an example of population consisting of four types and 5 stable states.



We are interested in which of the $m + 1$ equilibria the process will spend most of its time. We investigate this issue in the next section by analysing the process under the limit of stochastic stability.

³⁰The remaining transition probabilities are given in the appendix.

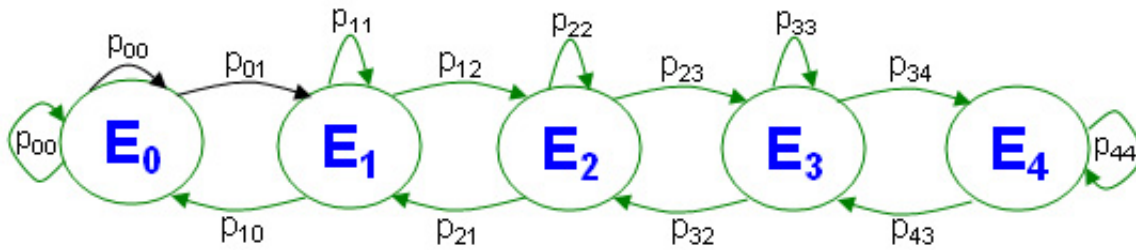
3.3 Stochastic Stability

In order to find clear conditions that determine which equilibria the process will predominantly be in we allow the mutation rate of the process to vanish to zero. Under this familiar limit we are able to determine the equilibria that the process will be located in the long run, as detailed below.³¹

Let the time spent in equilibrium E_i in the long run stationary distribution be given by π_i . Then an equilibrium E_i is defined as being *stochastically stable* if

$$\lim_{\varepsilon \rightarrow 0} \pi_i > 0$$

Under vanishing noise only the very minimum escape jump is of importance, jumps to states other than the neighbours of an equilibria quickly become negligible as ε decreases to 0. As such under stochastic stability the $m + 1$ Markov chain can be seen as a birth and death Markov chain. An example of such a chain is shown below for a four type population, generating 5 stable states.



When determining which of two neighbouring equilibria receives more weight as ε vanishes, the size of the relevant basins are the only factor which determine selection. Any combinational factors are overwhelmed by the strength of the vanishing mutation rate. Similarly the issue of

³¹There may be more than one equilibria.

possible counter mutations plays no part in the analysis.

And so we are able to determine that if an equilibria requires more mutations to pass its neighbour's threshold than vice versa, then it will receive more weight in the long run. By comparing the stochastic nature of neighbouring basin jumps throughout the $m + 1$ state space we are able to produce the exact conditions that describe which equilibria will be stochastically stable as detailed in proposition 1.

Proposition 1

An equilibria, E_k will be stochastically stable, $\lim_{\epsilon \rightarrow 0} \pi_k > 0$, if and only if for every $j \in \{0, \dots, m - 1\}$

$$S_j^k > 0$$

$$\begin{cases} S_j^k = \sum_{r=k}^{r=j} 2(p_{r+1} - \sum_{s=0}^{s=r} \alpha_s) - \alpha_{r+1} & \text{For } m > j \geq k \\ S_j^k = \sum_{r=j}^{r=k-1} \alpha_{r+1} - 2(p_{r+1} - \sum_{s=0}^{s=r} \alpha_s) & \text{For } j < k \end{cases}$$

Proof. See Appendix. ■

As in my previous paper³², under vanishing noise we are dealing with basin size as the only factor determining the stochastic stability of the process. Interestingly however the general heterogeneity of the current model creates several linked basins of attractions. In this paper I shall describe 'flow' from basins A to B as a larger probability of jumping to A from B than vice versa and 'steepness' to be the magnitude of this difference. Between neighbours, the direction and steepness of the flow is found purely by comparing the size of the respective basin jumps. Each basin jump is determined from the proportions of the relevant types and the strength of their preferences.

In order for an equilibria, E_k , to be stochastically stable firstly there

³²Noise Matters in Heterogenous Populations.

must be flow towards it from its immediate neighbours, E_{k-1} and E_{k+1} . Next, if there exists positive flow from its neighbours towards their other neighbour, E_{k-2} and E_{k+2} , this flow cannot be larger than the flow E_k receives from E_{k-1} and E_{k+1} . And this combined steepness cannot then be less than any flow that may exist from E_{k+2} towards neighbour E_{k+3} . And so on. If and only if all these requirements are made can equilibria E_k be stochastically stable.

Hence the values of p_i and α_i are critical in determining not just whether B_i is stochastically stable or not, but they may also have a large impact on which of the other m states in the process as a whole will be stochastically stable.

We now look to apply proposition 1 to a specific form of the model, in order to deduce the characteristics of the stochastically stable equilibrium for an interesting case of a population of three types.

3.4 The Influence of a Neutral Group in a Three Type Population

In this section we investigate the long-run location of a population consisting of three types of player. In particular we wish to consider a population that initially had two types of players with opposing preferences, and then at a later date experienced a group of relatively neutral players enter the population.³³ We look to examine what effect the existence of a neutral group can have on the population as a whole and what sway any small preferences of the neutral group away from indifference may have on a society. We also look to investigate whether the neutral group can increase the welfare of a previously divided society.

There are many interesting possible applications of such a model where a stalemate exists between two types of players, often to the detriment of society. For example consider a population requiring a significant majority vote in order for change to occur. Indeed one could consider long term conflict in Northern Ireland caused in part by strong opposing views between catholic and protestant residents. Here a long term co-existence of viewpoints could possibly be resolved by a small group of neutral types entering the population.³⁴ Other conflicts such as the Gaza strip crisis could also be used as an example where the possibility of a neutral group swaying both societies towards agreement could be of interest.

There are also technological applications of the model, for example the recent growth of networking websites such as Facebook, MySpace and Bebo. Here people generally benefit from more of their friends being on

³³One could also consider over time some proportion of each type changing their mind and becoming less extreme in their views.

³⁴Or a new generation growing up with more neutral views, or people changing their minds.

their website of choice, yet there is evidence that certain types of people seem to prefer different sites to others, possibly preventing one website to dominate so far.³⁵ For instance it is believed that socio-economics plays a large role in people's network preferences, for instance one researcher has found "students whose parents have less than a high school education are more likely to be MySpace users, while students whose parents have a college education are more likely to be Facebook users than others". As more people look to join a network site, probably of fairly neutral opinions on which to join, it could be of interest whether this will lead to one website dominating the whole market.³⁶

A further application could be education in classrooms. In some schools the behavior of students is a large issue within the classroom. One could consider a group of students who are focused on learning, and another opposing group who have a preference for disruptive behavior. Here a group of neutral students could be very influential in the long term behavior, culture and success of the class.

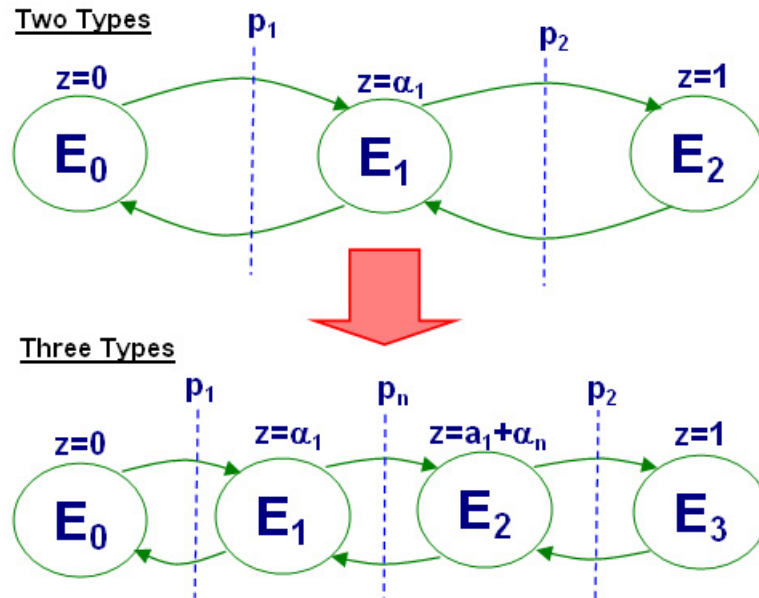
We now go on to analyse the influence of such a neutral group in a three type population.

³⁵At the time of writing Myspace and Facebook hold the majority of the market in roughly equal proportions.

³⁶<http://crookedtimber.org/2007/11/21/using-facebook-vs-myspace/> .

3.4.1 The Analysis

To begin with let us consider a population with three types of players, $\{T_1, T_2, T_3\}$, creating four equilibria, $\{E_0, E_1, E_2, E_3\}$, of which E_0 and E_3 are monomorphic, and E_1 and E_2 are states of co-existence. However for the convenience of this section, let us relabel the types as $\{T_1, T_N, T_2\}$ ³⁷ considering that T_1 agents have a preference to co-ordinate on s_1 , while the now T_2 agents prefer to co-ordinate on s_2 , and T_N agents have no strong preference and we will describe them as *neutral*. The proportion of each type within the population is represented by $\{\alpha_1, \alpha_N, \alpha_2\}$ with thresholds $p_1 < 0.5$, p_n and $p_2 > 0.5$ respectively. We make the assumption T_1 and T_2 agents are equally represented at all times such that $\alpha_1 = \alpha_2 = \frac{1-\alpha_N}{2}$. We represent the situation in the diagram below³⁸



³⁷Such that from section 1 T_1 agents are denoted by T_1 , T_2 agents are denoted by T_N and T_3 agents are denoted by T_2 .

³⁸ p_1 is drawn relatively closer to α_1 than p_2 is to E_2 as we will at times later consider type 1 agents having relatively stronger preferences than type 1.

Given the model above we are able to apply proposition 1 and determine three conditions for E_0 to be stochastically stable.³⁹

Lemma 1 $\lim_{\varepsilon \rightarrow 0} \pi_0 > 0$ if conditions 1a, 2a and 3a are satisfied, where

$$\begin{aligned} 1a) \quad p_1 &\geq \frac{1 - \alpha_n}{4} \\ 2a) \quad p_n &\geq \frac{3 - \alpha_n}{4} - p_1 \\ 3a) \quad p_n &\geq \frac{3}{2} - (p_1 + p_2) \end{aligned}$$

Proof. Please see appendix. ■

Condition 1a details that a positive flow from E_1 to E_0 is required for the stochastic stability of E_0 . Condition 2a details that the combined balance of flow from basins B_0 and B_1 must be towards E_0 , and condition 3a represents that the combined flow from all three basins must be towards E_0 . Interestingly unless conditions 1a and 2a are fulfilled, type 2 agent's preferences have no influence on whether E_0 will be selected.

We are able to make the conditions of Lemma 1 more succinct and clarify the influence of the neutral types preferences. The nature of type 2 agent's preferences, p_2 , dictate whether satisfying condition 2a implies that condition 3a is simultaneously satisfied. Given that condition 1a is satisfied, we can reduce the conditions for E_0 to be stochastically stable to just one more condition.

Lemma 2 For any given p_1, p_n, p_2 and α_n

$\lim_{\varepsilon \rightarrow 0} \pi_0 = 1$ if and only if

$$\begin{aligned} p_1 &> \frac{1 - \alpha_n}{4} \text{ and} \\ \begin{cases} p_n > \frac{3 - \alpha_n}{4} - p_1 & \text{if } p_2 > \frac{3 + \alpha_n}{4} \\ p_n > \frac{3}{2} - (p_1 + p_2) & \text{if } p_2 < \frac{3 + \alpha_n}{4} \end{cases} \end{aligned}$$

³⁹We focus on the stochastic stability of E_0 for clarity of explanation. The conditions for stochastic stability of E_1 , E_2 and E_3 are given in the appendix.

Proof. Please see appendix. ■

Consider that $p_1 > \frac{1-\alpha_n}{4}$ such that there is positive flow from E_1 to E_0 . Then if $p_2 > \frac{3+\alpha_n}{4}$ there is positive flow from E_3 to E_2 , a combination of type 1 and the neutral group's preferences determine whether there is positive flow from E_2 towards E_0 , which is required for E_0 to be stochastically stable. If $p_2 > \frac{3+\alpha_n}{4}$ then there is flow from E_1 to E_0 but also positive flow from E_2 to E_3 , away from E_0 , and here the a combination of the strength of all three types preferences determines equilibrium selection.

We now look at a specific form of a co-ordination game, and apply the conditions for stochastic stability above to a variety of settings.

3.4.2 Pairwise Matching

In this section we consider pairwise matching where an agent has an equal chance of playing a stage game with any other agent in the population each period.

We can now analyse which particular equilibria will be stochastically stable under a variety of parameters. There are two main issues of interest which we wish to investigate. The first issue is how small deviations from indifference in the neutral group's preferences may effect the equilibrium selection of the whole population. This is achieved by assessing the influence of a neutral group's preferences under varying strengths of heterogeneity between type 1 and 2 agents. The second issue is whether the very introduction of a small neutral group itself can change equilibrium selection.

To begin let us describe the various stage games between agent types.

The stage game between a T_1 and T_2 is given by:⁴⁰

T_1, T_2	s_1	s_2
s_1	$\theta, 1$	$0, 0$
s_2	$0, 0$	$1, k\theta$

where $\theta > \max\{\frac{1+\alpha_n}{k(1-\alpha_n)}, \frac{1+\alpha_n}{1-\alpha_n}\}$.⁴¹ The level of θ represents the level of heterogeneity between the two agents. We allow $k > 0$ to represent the possibility of a certain type possessing relatively stronger preferences for their preferred strategy, $k = 1$ being equivalent to Type 1 and 2 agents having equal but opposite preferences.

Similarly let stage game between two T_n agents be:⁴²

T_N, T_N	s_1	s_2
s_1	$\frac{\theta}{2}$	0
s_2	0	$\frac{q\theta}{2}$

where q is constant and $q = 1$ corresponds with perfectly neutral agents.⁴³ Variations in q will allow us to consider the effects of small preferences the neutral group may possess. We obtain the thresholds $p_1 = \frac{1}{1+\theta}$, $p_n = \frac{1}{1+q}$ and $p_2 = \frac{k\theta}{1+k\theta}$ from the stage games, detailing the proportion of agents required to play s_1 in the population in order for all T_i agents to have s_1 as their best response.

Before continuing we wish to discuss social welfare in different equilibria. Total social welfare is often at its lowest in the co-existence equilibria as the conflict between the type one and two agents significantly reduces the

⁴⁰From section 1, this is equivalent to $\rho = 1$, $\beta_1 = \theta$, $\gamma_1 = 1$, $\beta_2 = 1$, $\gamma_1 = \theta$ and $\gamma_2 = k\theta$.

⁴¹This ensures that $p_k < \sum_{i=1}^{i=k} \alpha_i < p_{k+1}$ and $k\theta > 1$.

⁴²The 4 remaining stage games are not given here but are easily obtained.

⁴³The payoff $\frac{\theta}{2}$ has been chosen arbitrary, and will only have an effect when considering welfare.

network effect, the neutral group suffering the most from the lack of coordination. Only if one group's preferences are extreme will a co-existence equilibrium not be the worst place for the population as a whole to be in.

Lemma 3

A) For $k \simeq 1$, the total social welfare at E_1 or E_2 is less than that of either E_0 or E_3 for any value of α_n , j and θ ,

and

B) For any value of k , E_1 and E_2 can never be the equilibria which maximizes total social welfare for any value of α_n , j and θ .

Proof. Given in the Appendix. ■

We now go onto the analysis of equilibrium selection in our three type population.

3.4.3 Equilibrium Selection Results

In this section we investigate the equilibrium selection for general parameters θ , α_n , p_n and k . In doing so we also look to examine the impact of the neutral group's preferences on the population as a whole. We are interested in the neutral group's influence for varying degrees of heterogeneity between agents one and two.

We find that the preferences of the neutral group can indeed sway equilibrium selection, and their influence is highly dependant on the heterogeneity of the two opposing types.

Proposition 2 characterizes the equilibrium selection results for our general three type model.

Proposition 2 Consider $k \geq 1$ and a population of neutral types α_n with a threshold level of p_n . Then:⁴⁴

For $\theta < \frac{3+\alpha_n}{k(1-\alpha_n)}$, the stochastically stable equilibrium will be

$$\begin{cases} E_0 & \text{if } p_n > \frac{3}{2} - \frac{1+2k\theta+k\theta^2}{(1+\theta)(1+k\theta)} \\ E_3 & \text{if } p_n < \frac{3}{2} - \frac{1+2k\theta+k\theta^2}{(1+\theta)(1+k\theta)} \end{cases}$$

For $\frac{3+\alpha_n}{k(1-\alpha_n)} < \theta < \frac{3+\alpha_n}{(1-\alpha_n)}$, the stochastically stable equilibrium will be

$$\begin{cases} E_0 & \text{if } p_n > \frac{3-\alpha_n}{4} - \frac{1}{1+\theta} \\ E_2 & \text{if } p_n < \frac{3-\alpha_n}{4} - \frac{1}{1+\theta} \end{cases}$$

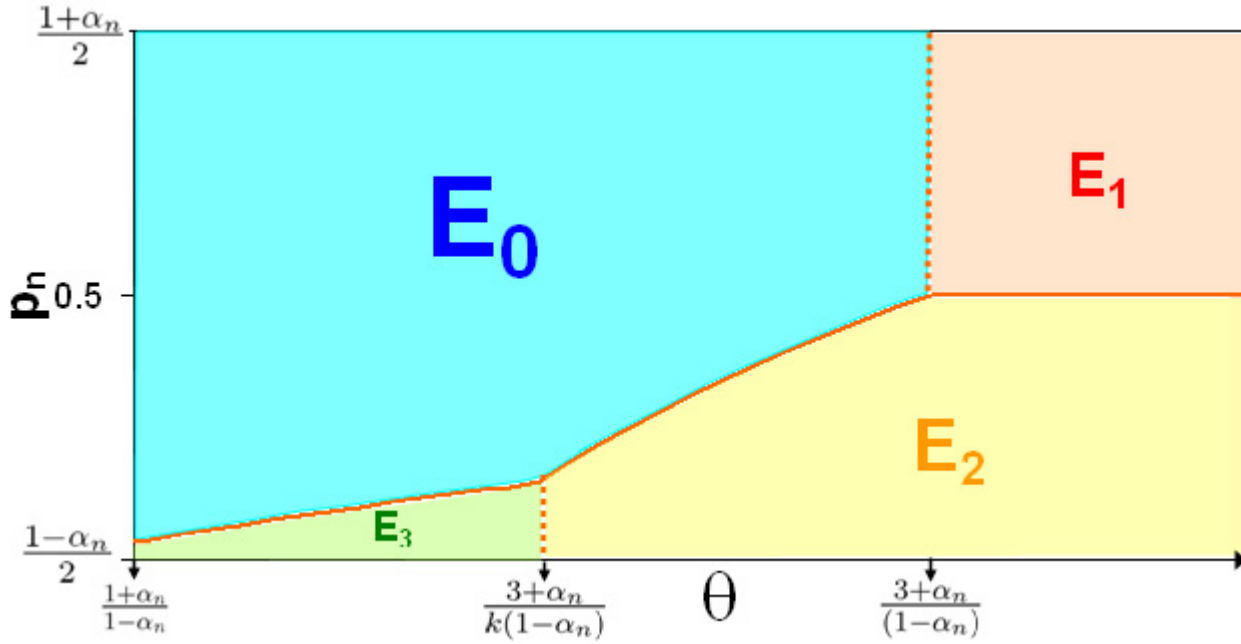
For $\theta > \frac{3+\alpha_n}{(1-\alpha_n)}$, the stochastically stable equilibrium will be

$$\begin{cases} E_1 & \text{if } p_n > \frac{1}{2} \\ E_2 & \text{if } p_n < \frac{1}{2} \end{cases}$$

Proof. See appendix ■.

⁴⁴We use p_n rather than $\frac{1}{1+q}$ for ease of explanation. $k < 1$ results are symmetric.

Proposition 2 is represented by the diagram below, with specific values of $\alpha_n = 0.2$ and $k = 1.5$.



When a low degree of heterogeneity exists, the neutral group's preferences decide between which of the monomorphic equilibria will be selected. As $k > 1$ here, the neutral group must have a preference towards strategy 1 in order for E_3 to be stochastically stable. The weak level of heterogeneity means that neither of the two co-existence equilibria will be a long term equilibrium, but instead act as stepping stones between the two monomorphic equilibria.

However when the level of heterogeneity increases beyond a critical point it is possible for co-existence to be stochastically stable. Here the neutral's preferences can sway equilibrium selection between E_0 and the co-existence equilibrium E_2 , but it is now impossible for E_3 to be stochastically stable.

When the heterogeneity level increases further beyond a second critical point, given a stable size of the neutral group, the large degree of heterogeneity means that co-existence must now be the long term equilibrium of the population. Here the neutral group's preferences will only determine which of the two co-existence equilibria will be stochastically stable.

We now investigate what effect the very introduction of a perfectly neutral group can have on society.

3.4.4 The Effect of the Introduction of a Neutral Group

We now wish to analyse whether the introduction of neutral agents itself can effect the equilibrium selection of the whole population. We will see there is indeed a critical mass of the neutral group that will change equilibrium selection.

As the neutral group's size increases, the co-existence equilibria E_1 and E_2 are pushed away from each other and towards thresholds p_1 and p_2 respectively. Once both equilibria are close enough to these thresholds neither can be stochastically stable. The more extreme the opposing views of type one and two agents, the larger the neutral group must be in order to make a difference. Lemma 4 describes the relationship between the size of the neutral group and the equilibrium selected.

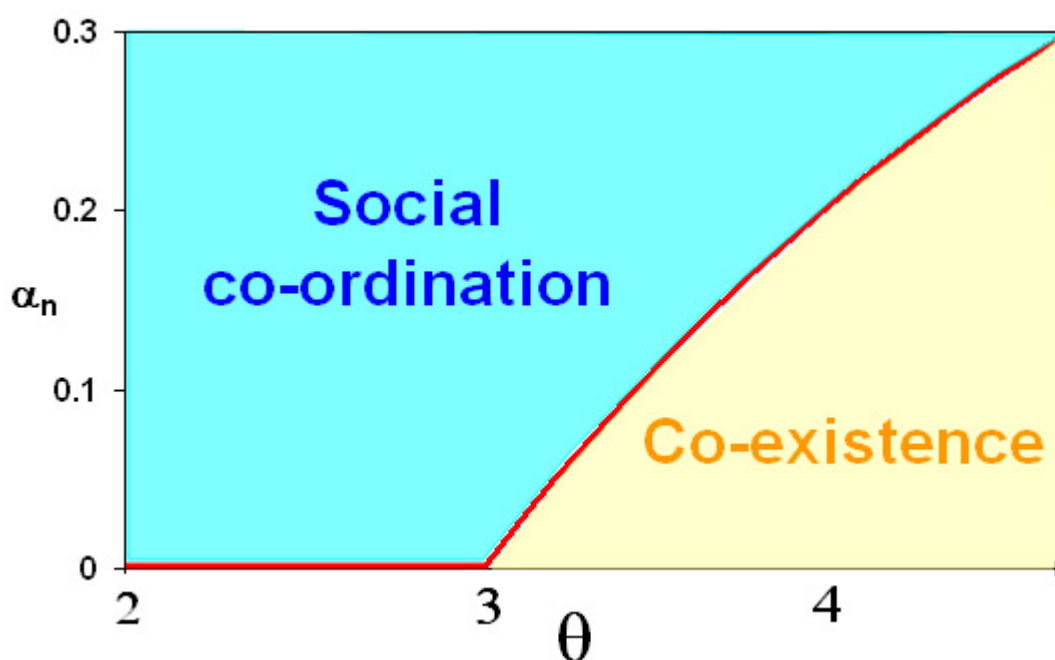
Lemma 4 $\lim_{\varepsilon \rightarrow 0} \pi_i = 0 \forall i \in \{1, 2\}$ if

$$\begin{cases} \alpha_n > \frac{\theta-3}{1+\theta} & \text{if } k \geq 1 \forall q \\ \alpha_n > \frac{\theta-3}{1+k\theta} & \text{if } k < 1 \forall q \end{cases}$$

There exists a level of $\alpha_N \in (\max(\frac{\theta-3}{1+\theta}, \frac{\theta-3}{1+k\theta}), 1]$ such that co-existence equilibria cannot be stochastically stable.

Proof. See Appendix. ■

And so the introduction of a neutral group of a large enough size ensures the long run equilibrium will not be one of co-existence. If the population were divided before the neutral group's entrance, then their existence alone could create a *welfare improving switch* in the population's long term location. The diagram below illustrates lemma 4 drawn for $k = 1$.



For $\alpha_n = 0$, we have a two type case along the horizontal axis in which co-existence will always prevail for $\theta > 3$.⁴⁵ However we can see that just a small proportion of neutrals in a population can be enough to upset co-existence when the opposing types are not too extreme. In such cases the co-existence equilibria now simply act as *stepping stones* between the two monomorphic equilibria.

⁴⁵This is the two type model of my previous paper NMHP with equal type 1 and two representations.

Bringing lemmas 3 and 4 together we can see that the existence of a neutral group can have a significant and positive effect on equilibrium selection and lead to a welfare improving switch.

Proposition 3 Consider a population where $\theta > 3$, $k > 1$ and $\alpha_n = 0$ such that we have a two type population where co-existence is stochastically stable. Then the introduction of a neutral group of size $\alpha_N > \frac{\theta-3}{1+\theta}$ ensures the social welfare maximizing equilibrium E_0 is now the stochastically stable selected equilibrium.

Proof. Proposition 3 follows from the results of Lemma 3 and 4. ■

And so the very introduction of a neutral population can have significant benefit to a society, the impact possibly being that in a previously divided society the socially optimum equilibrium is now selected. A society of purely neutral agents will naturally select the welfare maximising equilibria in a single type KMR model. However interestingly for some cases only a small amount of neutral agents are required to upset co-existence and switch the long run location to the welfare maximising equilibria.

3.5 Conclusion

This paper successfully characterises the stochastic stability of a generalised version of full player heterogeneity in KMR's original model. We then apply this characterisation to a specific three player type population. We can see that the entrance of a neutral group into a two type population can have two significant effects. Firstly, small preferences of the neutral group can significantly influence the equilibrium selection of the whole population. The force of this influence depending on the level of heterogeneity among the opposing groups. If this heterogeneity is extreme

a small group of neutral individuals can do little to influence the population. But if heterogeneity is less extreme a neutral group's views can have a very important effect in swaying a population towards social agreement. Secondly, the very introduction of a neutral group can result in co-existence no longer being the long run equilibrium of the population. Importantly in some cases only a small proportion of neutral agents are required to upset co-existence, and indeed this equilibrium switch can maximise the welfare of society as a whole.

3.6 Appendix

Proof of Proposition 1

Firstly, as we have a birth and death process, the long run time spent in each basin B_k can be found from the Markov stationary distribution and is given by

$$\pi_k = \frac{1}{1 + \sum_{r=0}^{r=k-1} \prod_{j=0}^{j=r} \frac{p_{k-j,k-j-1}}{p_{k-j-1,k-j}} + \sum_{r=0}^{r=m-k-1} \prod_{j=0}^{j=r} \frac{p_{k+j,k+j+1}}{p_{k+j+1,k+j}}}$$

We can describe the probability $p_{k,k+1}$ as follows

$$p_{k,k+1} = \varepsilon^{(p_{k+1})_+ - \sum_{i=1}^{i=k} \alpha_i} [f(1 - \varepsilon) + \rho f(\varepsilon)]$$

where ρ is some function independent of ε .

And similarly,

$$p_{k+1,k} = \varepsilon^{(\sum_{i=1}^{i=k+1} \alpha_i - [p_{k+1}]_-)} [f(1 - \varepsilon) + \rho f(\varepsilon)]$$

And so essentially,

$$\frac{p_{k,k+1}}{p_{k+1,k}} = \varepsilon^{2(p_{k+1} - \sum_{i=1}^{i=k} \alpha_i) - \alpha_{k+1}} [f(1 - \varepsilon) + \rho f(\varepsilon)]$$

and therefore,

$$\prod_{j=0}^{j=r} \frac{p_{k+j,k+j+1}}{p_{k+j+1,k+j}} = \varepsilon^{\sum_{r=j}^{r=m-1} 2(p_{r+1} - \sum_{i=1}^{i=r} \alpha_i) - \alpha_{r+1}} [f(1 - \varepsilon) + \rho f(\varepsilon)].$$

and

$$\prod_{j=0}^{j=r} \frac{p_{k+j,k+j+1}}{p_{k+j+1,k+j}} = \varepsilon^{\sum_{r=0}^{r=j} \alpha_{r+1} - 2(p_{r+1} - \sum_{i=1}^{i=r} \alpha_i)} [f(1 - \varepsilon) + \rho f(\varepsilon)].$$

For any value of r ,

$$\prod_{j=0}^{j=r} \frac{p_{k+j,k+j+1}}{p_{k+j+1,k+j}} \equiv \varepsilon^{s(r)} [f(1 - \varepsilon) + \rho f(\varepsilon)] \text{ and } \prod_{j=0}^{j=r} \frac{p_{k-j,k-j-1}}{p_{k-j-1,k-j}} \equiv \varepsilon^{t(r)} [f(1 - \varepsilon) + \rho f(\varepsilon)].$$

Then if for some r , $s(r) < 0$ or $t(r) < 0$,

$$\text{if } \lim_{\varepsilon \rightarrow 0} \prod_{j=0}^{j=r} \frac{p_{k+j,k+j+1}}{p_{k+j+1,k+j}} \rightarrow \infty \text{ or } \lim_{\varepsilon \rightarrow 0} \prod_{j=0}^{j=r} \frac{p_{k-j,k-j-1}}{p_{k-j-1,k-j}} \rightarrow \infty,$$

then $\lim_{\varepsilon \rightarrow 0} \pi_k = 0$.

As such an equilibria, E_k will be stochastically stable, $\lim_{\varepsilon \rightarrow 0} \pi_k > 0$, if and only if for every $j \in \{0, \dots, m-1\}$

$$S_j^k > 0$$

where

$$\begin{cases} S_j^k = \sum_{r=k}^{r=j} 2(p_{r+1} - \sum_{s=0}^{s=r} \alpha_s) - \alpha_{r+1} & \text{For } m > j \geq k \\ S_j^k = \sum_{r=j}^{r=k-1} \alpha_{r+1} - 2(p_{r+1} - \sum_{s=0}^{s=r} \alpha_s) & \text{For } j < k \end{cases}$$

■

Proof of Lemma 1 Condition 1a) is obtained from

$$\begin{aligned} S_0^0 &= 2(p_1 - 0) - \frac{1 - \alpha_n}{2} > 0 \\ \Rightarrow p_1 &> \frac{1 - \alpha_n}{4} \end{aligned}$$

Condition 1b) is obtained from

$$\begin{aligned} S_0^1 &= 2(p_1 - 0) - \frac{1 - \alpha_n}{2} + 2(p_n - \frac{1 - \alpha_n}{2}) - \alpha_n > 0 \\ \Rightarrow 2p_1 + 2p_n - \frac{3 - \alpha_n}{2} &> 0 \\ \Rightarrow p_n &> \frac{3 - \alpha_n}{4} - p_1 \end{aligned}$$

Condition 1c) is obtained from

$$\begin{aligned} S_0^2 &= 2(p_1 - 0) - \frac{1 - \alpha_n}{2} + 2(p_n - \frac{1 - \alpha_n}{2}) - \alpha_n + 2(p_2 - \frac{1 + \alpha_n}{2}) - \frac{1 - \alpha_n}{2} > \\ \Rightarrow 2p_1 + 2p_n + 2p_2 - \frac{3 - \alpha_n}{2} - \frac{3}{2} - \frac{\alpha_n}{2} &> 0 \\ \Rightarrow p_n &> \frac{3}{2} - (p_1 + p_2) \end{aligned}$$

■

Stochastic Stability of E_1 , E_2 and E_3

E_3 :

$\lim_{\varepsilon \rightarrow 0} \pi_3 = 1$ if conditions 1b, 2b and 3b are all satisfied, where

$$\begin{aligned} 1b) \quad p_2 &< \frac{3 + \alpha_n}{4} \\ 2b) \quad p_n &< \frac{5 + \alpha_n}{4} - p_2 \\ 3b) \quad p_n &< \frac{3}{2} - (p_1 + p_2) \end{aligned}$$

E_1 :

$\lim_{\varepsilon \rightarrow 0} \pi_1 = 1$ if conditions 1c, 2c and 3c are all satisfied, where

$$\begin{aligned} 1c) \quad p_1 &< \frac{1 - \alpha_n}{4} \\ 2c) \quad p_n &> \frac{1}{2} \\ 3c) \quad p_n &> p_2 - \frac{5 + \alpha_n}{4} \end{aligned}$$

E_2 :

$\lim_{\varepsilon \rightarrow 0} \pi_2 = 1$ if conditions 1d, 2d and 3d are satisfied, where

$$\begin{aligned} 1d) \quad p_2 &> \frac{3 + \alpha_n}{4} \\ 2d) \quad p_n &< \frac{1}{2} \\ 3d) \quad p_n &> p_1 - \frac{3 - \alpha_n}{4} \end{aligned}$$

Proof of Lemma 2

To prove lemma 2 consider that satisfying condition 2a also implies condition 3a is satisfied if and only if

$$\begin{aligned} \frac{3 - \alpha_n}{4} - p_1 &> \frac{3}{2} - (p_1 + p_2) \\ \Rightarrow p_2 &> \frac{3 + \alpha_n}{4} \end{aligned}$$

■

Proof of Proposition 2

Let us first note that from lemma 2 that as

$$p_1 > \frac{1-\alpha_n}{4} \Rightarrow \frac{1}{1+\theta} > \frac{1-\alpha_n}{4} \Rightarrow \theta < \frac{3+\alpha_n}{(1-\alpha_n)}$$

and

$$p_2 < \frac{3+\alpha_n}{4} \Rightarrow \frac{k\theta}{1+k\theta} < \frac{3+\alpha_n}{4} \Rightarrow \theta < \frac{3+\alpha_n}{k(1-\alpha_n)}$$
 we can express lemma 2 as:

For any given p_1, p_n, p_2 and α_n

$\lim_{\varepsilon \rightarrow 0} \pi_0 > 0$ if and only if

$$\theta < \frac{3+\alpha_n}{(1-\alpha_n)} \text{ and}$$

$$\begin{cases} p_n > \frac{3-\alpha}{4} - p_1 & \text{if } \theta > \frac{3+\alpha_n}{k(1-\alpha_n)} \\ p_n > \frac{3}{2} - (p_1 + p_2) & \text{if } \theta < \frac{3+\alpha_n}{k(1-\alpha_n)} \end{cases}$$

Therefore we can see that if $\theta < \frac{3+\alpha_n}{k(1-\alpha_n)}$ then both conditions 1a and 1b are satisfied (1c and 1d not satisfied) such that E_1 and E_2 can *not* be stochastically stable. Then from lemma 2 and condition 3c we can determine the stochastically stable equilibrium will be

$$\begin{cases} E_0 & \text{if } p_n > \frac{3}{2} - \frac{1+2k\theta+k\theta^2}{(1+\theta)(1+k\theta)} \\ E_3 & \text{if } p_n < \frac{3}{2} - \frac{1+2k\theta+k\theta^2}{(1+\theta)(1+k\theta)} \end{cases}$$

Now consider that $\frac{3+\alpha_n}{(1-\alpha_n)} < \theta < \frac{3+\alpha_n}{k(1-\alpha_n)}$. Condition 1d is *now* satisfied (1b not satisfied) so that E_3 cannot be stochastically stable, however as $k > 1$ condition 1a is still satisfied such that E_0 can be stochastically stable and E_1 cannot, and from lemma 2 and condition 2d we can determine the stochastically stable equilibrium will be,

$$\begin{cases} E_0 & \text{if } p_n > \frac{3-\alpha_n}{4} - \frac{1}{1+\theta} \\ E_2 & \text{if } p_n < \frac{3-\alpha_n}{4} - \frac{1}{1+\theta} \end{cases}$$

Finally consider that $\theta > \frac{3+\alpha_n}{(1-\alpha_n)}$. Then conditions 1a and 1b are not satisfied, and so neither E_0 or E_3 can be stochastically stable. Therefore

from conditions 2c and 2d we can see the stochastically stable equilibrium will be as follows:

$$\begin{cases} E_1 & \text{if } p_n > \frac{1}{2} \\ E_2 & \text{if } p_n < \frac{1}{2} \end{cases}$$

■

Proof of lemma 3

A)

Essentially when considering moving from E_0 to E_1 type one agents gain $\alpha_1\theta - 1$ and type two agents lose $\alpha_1\theta$, therefore with neutral agents also having a decreased payoff at E_1 it is socially inferior to E_0 for k near to one.

More formally for $k = 1$ let us consider social welfare at E_0 for one period.

Type 1 agent's total payoff is given $\alpha_1 N$.

Type 2 agent's total payoff is given $\alpha_2\theta N = \alpha_1\theta N$ as type 1 and 2 agents are equally represented.

The neutral type's payoff is given by $\alpha_n j\theta N = \frac{1-\alpha_1}{2} j\theta$.

As such total social welfare at E_0 is given by

$$E_0^{tsw} = \alpha_1(1 + \theta) + \frac{1-\alpha_1}{2} j\theta$$

Now let us consider social welfare at the polymorphic equilibria E_1 .

Here type 1 agent's total payoff is given $\alpha_1^2\theta N$.

Type 2 agent's total payoff is given $(1 - \alpha_1)\alpha_1\theta$

The neutral type's payoff is given by $(1 - \alpha_1)\frac{1-\alpha_1}{2} j\theta$.

As such total social welfare at E_0 is given by

$$\begin{aligned} E_1^{tsw} &= \alpha_1^2\theta N + (1 - \alpha_1)\alpha_1\theta + (1 - \alpha_1)\frac{1-\alpha_1}{2} j\theta \\ &= \alpha_1\theta(\alpha_1 + (1 - \alpha_1)) + (1 - \alpha_1)\frac{1-\alpha_1}{2} j\theta. \end{aligned}$$

Therefore,

$$E_0^{tsw} - E_1^{tsw} = \alpha_1(1 - \alpha_1\theta) + \alpha_1(\alpha_1\theta) + \frac{1-\alpha_1}{2} j\theta$$

$$= \alpha_1 + \frac{1-\alpha_1}{2}j\theta > 0 \quad \forall j, \alpha_1.$$

Symmetrically $E_0^{tsw} > E_2^{tsw}$, $E_3^{tsw} > E_2^{tsw}$ and $E_3^{tsw} > E_2^{tsw}$. ■

B)

For $k > 1$ it is possible that the total social welfare at E_1 is greater than

E_3 as the total social welfare at E_1 is given by

$$E_1^{tsw} = \alpha_1\theta(\alpha_1 + (1 - \alpha_1)k) + (1 - \alpha_1)\frac{1-\alpha_1}{2}j\theta$$

and

$$E_3^{tsw} \text{ is given by } \alpha_1(1 + \theta) + \frac{1-\alpha_1}{2}j\theta$$

such that high levels of k yield $E_1^{tsw} > E_3^{tsw}$.

However in this case we have it that E_0 maximises social welfare not E_1

as

$$E_0^{tsw} = \alpha_1(1 + k\theta) + \frac{1-\alpha_1}{2}j\theta$$

such that

$$E_0^{tsw} - E_1^{tsw} = \alpha_1(1 - \alpha_1\theta) + \alpha_1(k\theta - (1 - \alpha_1)k\theta) + \alpha_1\frac{1-\alpha_1}{2}j\theta$$

$$= \alpha_1(1 - \alpha_1\theta) + \alpha_1(\alpha_1k\theta) + \alpha_1\frac{1-\alpha_1}{2}j\theta$$

$$= \alpha_1 + \alpha_1^2\theta(k - 1) + \alpha_1\frac{1-\alpha_1}{2}j\theta > 0 \quad \forall k > 1, \alpha_1, j.$$

Symmetrically for $k < 0$, $E_3^{tsw} > E_2^{tsw} \quad \forall \alpha_1, j$ and we see that the co-

existence equilibria can not maximise social welfare. ■

Proof of Lemma 4

From proposition 1 we can deduce that $\lim_{\varepsilon \rightarrow 0} \pi_1 = 0$ occurs when

$$S_0^1 = 2(p_1 - 0) - \frac{1-\alpha_n}{2} = 2\frac{\theta}{1+\theta} - \frac{1-\alpha_n}{2} > 0$$

$$\therefore \alpha_n > \frac{\theta-3}{1+\theta} \Rightarrow \lim_{\varepsilon \rightarrow 0} \pi_1 = 0.$$

Similarly, $\lim_{\varepsilon \rightarrow 0} \pi_2 = 0$ occurs when

$$S_2^3 = \frac{1-\alpha_n}{2} - 2(p_2 - \frac{1+\alpha_n}{2}) = \frac{1-\alpha_n}{2} - 2(\frac{k\theta}{1+k\theta} - \frac{1+\alpha_n}{2}) > 0$$

$$\therefore \alpha_n > \frac{\theta-3}{1+k\theta} \Rightarrow \lim_{\varepsilon \rightarrow 0} \pi_2 = 0.$$

and so

$\lim_{\varepsilon \rightarrow 0} \pi_i = 0 \forall i \in \{1, 2\}$ if

$$\begin{cases} \alpha_n > \frac{\theta-3}{1+\theta} & \text{if } k \geq 1 \forall q \\ \alpha_n > \frac{\theta-3}{1+k\theta} & \text{if } k < 1 \forall q \end{cases}$$

■

Transition Probabilities

The transition probability of escaping B_m and entering B_{m-1} in any period requires between $N - [qN]_-$ and $N - [pN]_+$ mutations and is given by

$$p_{m,m-1} = \sum_{i=N-[p_n N]_-}^{i=N-[p_{n-1} N]_+} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}$$

The probability of jumping from the polymorphic basin E_k to E_{k+1} is given by

$$p_{k,k+1} = \sum_{j=\sum_{i=0}^{i=k} \alpha_k N - [p_k N]_+}^{\sum_{i=0}^{i=k} \alpha_k N} \sum_{k=\max\{j - \sum_{i=0}^{i=k} \alpha_k N, 0\}}^{\min\{j, N-j\}} \left(\sum_{i=0}^{i=k} \alpha_k N - [p_k N]_- \right) \left((1 - \sum_{i=0}^{i=k} \alpha_k) N - k \right) \varepsilon^{\sum_{i=0}^{i=k} \alpha_k N + k - j} (1-\varepsilon)^{j-k} \varepsilon^k (1-\varepsilon)^{(1-\sum_{i=0}^{i=k} \alpha_k) N - k}$$

References

- Anwar, A. (1999) "On the Coexistence of Conventions (2002)". Journal of Economic Theory.
- Alan Beggs, (2002). "Large Deviations and Equilibrium Selection in Large Populations,". Economics Series Working Papers 129, University of Oxford, Department of Economics.
- Blume, L. (1999): "How Noise Matters" Working Paper.
- Bergin, J., and B. Lipman (1996): "Evolution with State-Dependent Mutations", *Econometrica*, 64, 943-956.

- Ellison, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047-71.
- (2000): "Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution" *Review of Economic Studies*, 67, 17-45.
- Granovetter, Mark. (1978). "Threshold Models of Collective Behavior." *American Journal of Sociology*, 83: 1420-1443.
- Hehenkamp B, (2001). "Equilibrium selection in the two-population KMR model". Discussion Papers in Economics, University of Dortmund, Department of Economics.
- Hahn, S. (1995): The Long Run Equilibrium in an Asymmetric Coordination Game mimeo, Harvard University.
- Hojman D, (2004) "Interaction Structure and the Evolution of Conventions."
- Kandori, Mailath and Rob(1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61, 29-56.
- Kolstad, I. (2003): "Viscosity and dispersion in an evolutionary model of learning."
- Myatt, D.P and C. Wallace (1998): "Adaptive Play by Idiosyncratic Agents," *Discussion Paper, Nuffield College, Oxford*.
- Norman, T. (2003a): "The Evolution of Co-ordination under Inertia."
- Robson, Arthur J. and Vega-Redondo, Fernando, (1996): "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *Journal of Economic Theory*, Elsevier, vol. 70(1), pages 65-92.
- Samuelson, I., (1994): "Stochastic Stability in Games with Alternative Best Replies," *Journal of Economic Theory* 64 , 35-65.
- Sandholm, W. H. (2006). "Pairwise comparison dynamics and evolution-

ary foundations for Nash equilibrium". Unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2007). "Population games and evolutionary dynamics". Unpublished manuscript, University of Wisconsin.

Schelling, Thomas C. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1: 143-186.

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

(1990) "Stochastic Evolutionary Game Dynamics" with Dean P. Foster. *Theoretical Population Biology*, 38, 219-232.

4 The Limitations of Stochastic Stability in Heterogenous Populations

Abstract

This paper seeks to assess the ability of stochastic stability results to represent populations subject to small but positive noise levels. Under varying positive noise levels we present numerical calculations of the long run distribution of heterogenous populations who play a binary choice co-ordination game. We show that populations with individual experimentation rates as *small* as one in a million periods can in fact be located in a completely different neighbourhood in the long run to the stochastic stability equilibrium. We also find that over a third of populations are misrepresented in this way by stochastic stability. As such we severely question the ability of vanishing noise analysis to represent the true nature of populations subject to even the smallest noise levels.

4.1 Introduction

This paper considers a model of the type introduced by Kandori, Mailath and Rob(1993, KMR) and Young(1993). KMRY have a single type of boundedly rational agents who occasionally experiment away from their best choice, giving the population movement and allowing the possibility of escaping two stable equilibria. In order to obtain clear selection results KMRY allow the experimentation rate to decrease to 0 which selects a single equilibria, described in the literature as the stochastically stable equilibrium.

In this paper we discuss a KMRY type model with a general level of heterogeneity such that we have several types of agents existing within a single population, each with different preferences in a binary choice coordination game. The general level of heterogeneity creates many stable equilibria and we study the equilibrium selection results of some specific populations.

The aim of the paper is to assess the validity of using stochastic stability as an equilibrium selection technique for a KMRY type model with a general level of heterogeneity in players' preferences. In particular we wish to investigate the ability of stochastic stability to represent the actual long term location of a population subject to *small but positive noise*.

The complex stochastic nature of the dynamics under positive noise entails that tractable results are difficult to produce, and in this paper we obtain our results instead through numerical calculations. By choosing specific populations we are able produce a discrete markov chain for a finite population size with set transition probabilities. As such we can calculate the exact location of the population in the long run for varying levels of noise, and compare these results to the stochastically selected

equilibrium. We find populations with individual experimentation rates as *small* as one in a million periods can in fact be located in a completely different neighbourhood in the long run to the stochastic stability equilibrium.

We also assess the likelihood that a population will have different equilibrium selection results under stochastic stability and small noise levels. We find that such a misrepresentation is not a rare event. Populations with greater degrees of heterogeneity are more likely to be misrepresented by stochastic stability. Indeed we find that more than a third of populations with five player types are misrepresented by stochastic stability.

This paper is not the first to look at KMRY type models with positive noise levels. Theoretical papers of Benaim and Weibull (2003a,b) and Binmore and Samuelson(1997) have constant positive noise levels in a population whose size is taken to infinity. Myatt and Wallace(1998) and Beggs(2002) devote some attention to positive noise levels. Sandholm(2005) shows in some homogenous population games constant noise results can differ from those of vanishing noise. My previous chapter 'Noise Matters in Heterogenous Populations' proves that in a two type population positive noise results can differ from those of stochastic stability. This paper continues by showing more heterogenous populations under extremely small noise levels can be located in very different neighborhoods to where stochastic stability would suggest.

The paper is presented as follows. The next chapter presents the model and the dynamics of the population. We then discuss our methods of producing the numerical calculations. In section 4 we present an introductory calculation, and then go onto demonstrate our key result for three separate populations under an array of small noise levels. Section 6 de-

tails the frequency of our results from section 4. Section 7 discusses the reasons behind our result and section 8 concludes.

4.2 The Model

Consider a single population of N players consisting of m different types of players, type i being denoted by T_i , $i \in \{1, \dots, m\}$. The players are in a co-ordination game as in any period the more players in the population playing an agent's current strategy, the higher the agent's payoff. However the different types of players within the population receive different payoffs each period, due to their personal preferences. Each player has two possible strategies to choose from.

The payoff in any period t for a T_i agent playing strategy s_k , $\lambda_i^{s_k}$, is given by

$$\begin{aligned}\lambda_i^{s_1} &= \beta_i(z(t) - 1)^\rho \\ \lambda_i^{s_2} &= \gamma_i(N - z(t) - 1)^\rho\end{aligned}$$

where $z(t)$ represents the number of agents playing s_1 in period t and where $\rho \in \mathbb{R}^+$.⁴⁶

We have it that $\beta_i > \beta_j \forall i < j$. As such T_1 players have the strongest preference to co-ordinate on s_1 , followed by T_2 players, and so on. Indeed, T_1 players may require only a small proportion of the population to play s_1 in order for s_1 to be their best response. Similarly, $\gamma_i < \gamma_j \forall i < j$ indicates T_m players have the strongest preference to co-ordinate on s_2 .⁴⁷

⁴⁶For most applications we would have $\rho \in (0, 1]$ but we leave $\rho > 1$ open for generality.

⁴⁷We leave $\beta_1 < \gamma_1$ open as a possibility for generality.

4.2.1 The Dynamics of a Heterogenous Population

Our population consists of myopic players who consider the state of play will remain the same as the previous period, $z(t-1)$. Therefore the best response of any T_i agent in any period t is⁴⁸

$$\begin{cases} s_1 & \text{if } z(t-1) > \frac{1}{1+(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}}} N + \frac{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} - 1}{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} + 1} \equiv p_i N \\ s_2 & \text{Otherwise} \end{cases}$$

We shall continue by denoting N_i as the number of T_i agents in a population, such that $N = \sum_{i=1}^{i=m} N_i$. We define the proportion of T_i agents in the population as $\alpha_i = N_i/N$.⁴⁹ As agents consider only the aggregate amount of players choosing a strategy, $z(t) = \{0, 1, \dots, N\}$ defines the state of the process at any time t .

Two monomorphic stable points for the process exist, E_m , where all agents choose to play $s_1(z = N)$ and E_0 where all choose $s_2(z = 0)$.⁵⁰

Furthermore, given

$$p_k < \sum_{i=1}^{i=k} \alpha_i < p_{k+1}$$

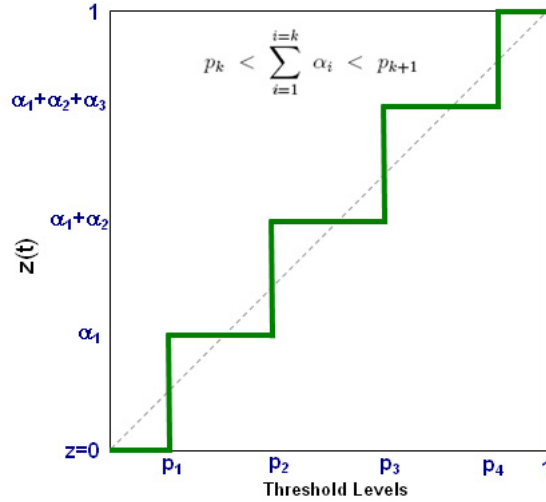
there exists $m-1$ stable points of co-existence denoted by $E_k, k \in \{1, \dots, m-1\}$ where $\sum_{i=1}^{i=k} \alpha_i$ agents will play s_1 and $\sum_{i=k+1}^{i=m} \alpha_i$ agents play s_2 . For the remainder of this paper we shall only consider cases where this constraint holds for all values of k .

⁴⁸Such that $p_i = \frac{1}{1+(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}}} + \frac{(\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} - 1}{((\frac{\beta_i}{\gamma_i})^{\frac{1}{\rho}} + 1)N}$.

⁴⁹We define $N_0 = 0$ and $\alpha_0 = 0$ for later analysis.

⁵⁰A notation change from the author's previous Noise Matters in Heterogenous Populations paper.

The constraint is represented by the diagram below which shows the proportion of a four type population who will play s_1 , given a proportion p_j played s_1 last period.



Overall there exists $m + 1$ possible stable states of the process, two of which are monomorphic, and $m - 1$ which are states of co-existence.

Let us continue by addressing the basins of attraction of each stable point and let the basin of attraction of E_k be denoted by B_k .

B_k is then defined as any state $z(t)$ in ⁵¹

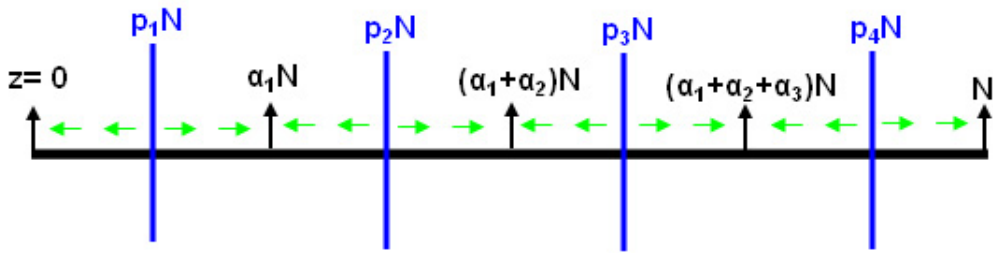
$$\begin{cases} \{0, \dots, [p_1 N]_-\} & \text{for } k = 0 \\ \{[p_m N]_+, \dots, N\} & \text{for } k = m \\ \{[p_k N]_+, \dots, [p_{k+1} N]_-\} & \text{for } k \in \{1, \dots, m-1\} \end{cases}$$

For a population of agents who always play a myopic best response the long run location of the process depends upon the initial set up alone. However, as in KMRY, any agent can select a strategy other than its best response with probability ε each period, I shall describe ε as the mutation

⁵¹Defining $[x]_-$ as the nearest integer below or equal to x and $[x]_+$ as the nearest integer above or equal to x .

experimentation rate.⁵² As it is possible to jump from any given state to any other state in one period, including itself, the process is irreducible and aperiodic, and therefore the markov chain is ergodic.

A state space using an example of population consisting of four types and 5 stable states is shown below



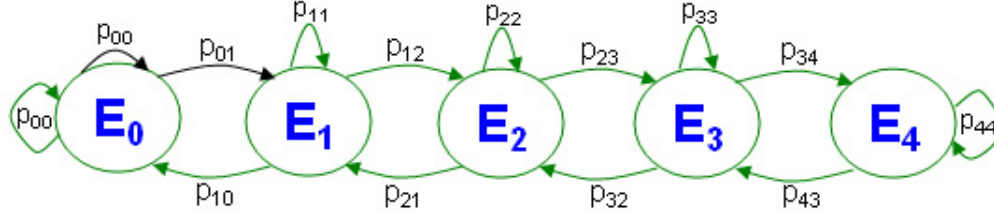
4.3 A Positive Level of Noise

As we will be dealing with positive levels of noise, the process will be in continual movement and therefore will not converge to a single point. However under the best reply dynamics of the population, at any state in a certain basin of attraction every agent of a particular type has the same best response next period. Therefore the number of mutations required to leave a basin, and the probability of this occurring, is the same for any state within a basin of attraction.

As such we are able to divide the state space into neighborhoods defined by the basins of attraction, and the state space $z(t) = \{0, 1, \dots, N\}$ can be reduced to an $m + 1$ state markov chain. We can then determine the proportion of time the process will spend in each of the neighbourhoods for various ε levels. It is possible to jump from one basin to any other each period, however the probability of the process jumping beyond a

⁵²This system is often described as simultaneous revisions in the literature.

neighbouring basin is extremely small relatively, and although included in the calculations for clarity will not be illustrated. With this in mind we present a four type model with 5 stable states as an example below:



The markov transition probabilities correspond to the basin escape probabilities. The probability of escaping B_0 to its immediate neighbouring basin B_1 in any period is given by the binomial probability ⁵³

$$p_{01} = \sum_{i=[p_1 N]_+}^{[p_2 N]_-} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i}$$

Similarly $p_{m,m-1}$ is given by

$$p_{m,m-1} = \sum_{i=N-[p_m N]_-}^{i=N-[p_{m-1} N]_+} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i}$$

Escaping from polymorphic states is a more complicated affair as we are dealing with the net of two binomials as simultaneous mutations from different types can cancel each other out, and so the probability of moving from the polymorphic neighbourhood B_k to B_{k-1} is given by

$$p_{k,k-1} = \sum_{j=0}^{[p_k N]_-} \sum_{k=\max\{j-\sum_{i=0}^{i=k} \alpha_k N, 0\}}^{\min\{j, N-j\}} \left(\frac{\sum_{i=0}^{i=k} \alpha_k N}{\sum_{i=0}^{i=k} \alpha_k N - [p_k N]_-} \right) \left(\frac{(1 - \sum_{i=0}^{i=k} \alpha_k) N}{(1 - \sum_{i=0}^{i=k} \alpha_k) N - k} \right) \varepsilon^{\sum_{i=0}^{i=k} \alpha_k N + k - j} (1 - \varepsilon)^{j-k} \varepsilon^k (1 - \varepsilon)^{(1 - \sum_{i=0}^{i=k} \alpha_k) N - k}$$

We have a new ergodic markov chain whose long run behavior is given by the stationary equations $\pi^\varepsilon P^\varepsilon = \pi^\varepsilon$, where P^ε is the transition matrix

⁵³We only present the jumps p_{01} , $p_{m,m-1}$ and $p_{k,k-1}$ for establish the nature of the probability dynamics.

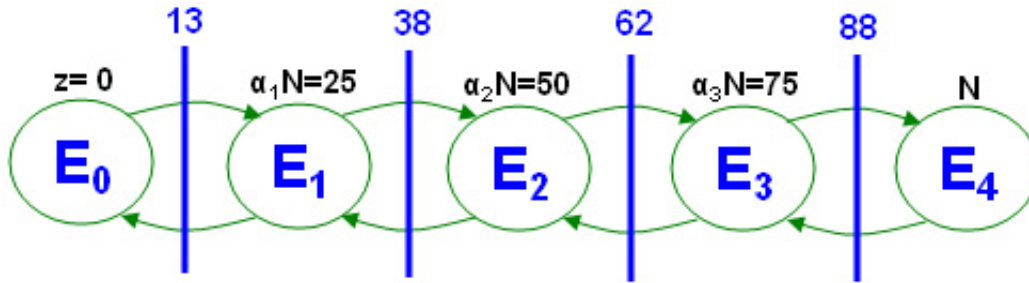
containing all transition probabilities of p_{ij} , and π^ε is the unique solution for fixed P^ε .

As such for a fixed population with specific player proportions α_i and preferences p_i , for a chosen level of ε we are able to calculate the stationary distribution of the process $\{\pi_0, \dots, \pi_m\}$, and determine the proportion of time the process will spend in each neighbourhood in the long run.

We proceed to look at a specific population and investigate whether variations in levels of ε , when ε is itself small, has any significant affect on the stationary distribution of the process.

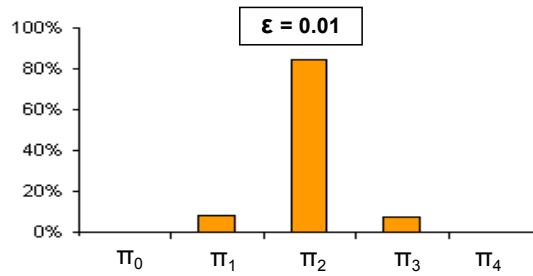
4.4 An Introduction to the Results

We begin by considering a specific population of 100 players consisting of four player types equally represented such that $\alpha_i = \frac{1}{4}$ for all $i \in \{1, 2, 3, 4\}$. Each player type has separate preferences yielding thresholds $p_1 = 0.13$, $p_2 = 0.38$, $p_3 = 0.62$ and $p_4 = 0.88$ for types one to four respectively. As such the state space can be viewed as



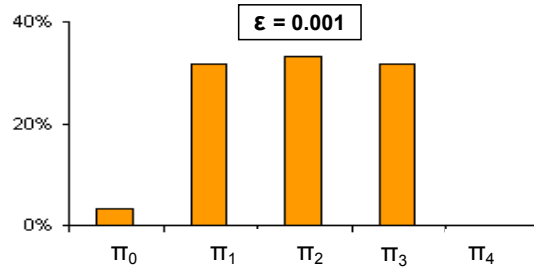
The aim of the paper is to assess the ability of stochastic stability analysis to represent the behaviour of populations subject to small but positive noise. And so we now look at the long run stationary distribution of the current process under three separate small values of the mutation rate to assess the consistency of the results.

We begin by calculating the stationary distribution under a positive noise level $\varepsilon = 0.01$, an individual mutation rate of one every one hundred periods. The results of the calculations are shown below.



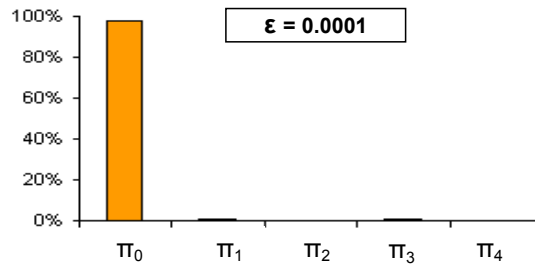
We find that for this small level of noise the process spends all of its time in states of co-existence, with the vast majority being in the middle equilibria, where half of the population play each strategy, and no time is spent in either monomorphic equilibria.

Next we reduce the mutation rate of the population to $\varepsilon = 0.001$ to assess the consistency of our first result, and the resulting stationary distribution is given below.



We find that the stationary distribution of the population for $\varepsilon = 0.001$ is significantly different from a mutation rate of $\varepsilon = 0.01$, despite both mutation levels being small. Here the stationary distribution details a balanced amount of time being spent in all three of the co-existence states, with a small amount of time being spent in B_0 .

Lastly we reduce the mutation rate further to $\varepsilon = 0.0001$ and the resulting stationary distribution is given below.



Surprisingly we find a *very* different distribution is produced when considering $\varepsilon = 0.001$ and $\varepsilon = 0.0001$ noise levels. Under $\varepsilon = 0.001$ the population will spend almost all of its time in B_0 and essentially no time in states of co-existence. Reducing ε to 0 we can verify that E_0 is the stochastically stable state, the equilibria where stochastic stability tells us the process *will spend all* of its time.⁵⁴

Stochastic stability informs us that this population will spend all of its time in the neighbourhood B_0 yet our numerical calculations show that with a mutation rate as small as $\varepsilon = 0.001$ this population will be mostly located in neighbourhoods of co-existence. For $\varepsilon = 0.01$ the long run location of the population is the opposite to where stochastic stability suggests. In the next section we show such results can indeed repeat themselves, and under exceptionally small levels of noise.⁵⁵

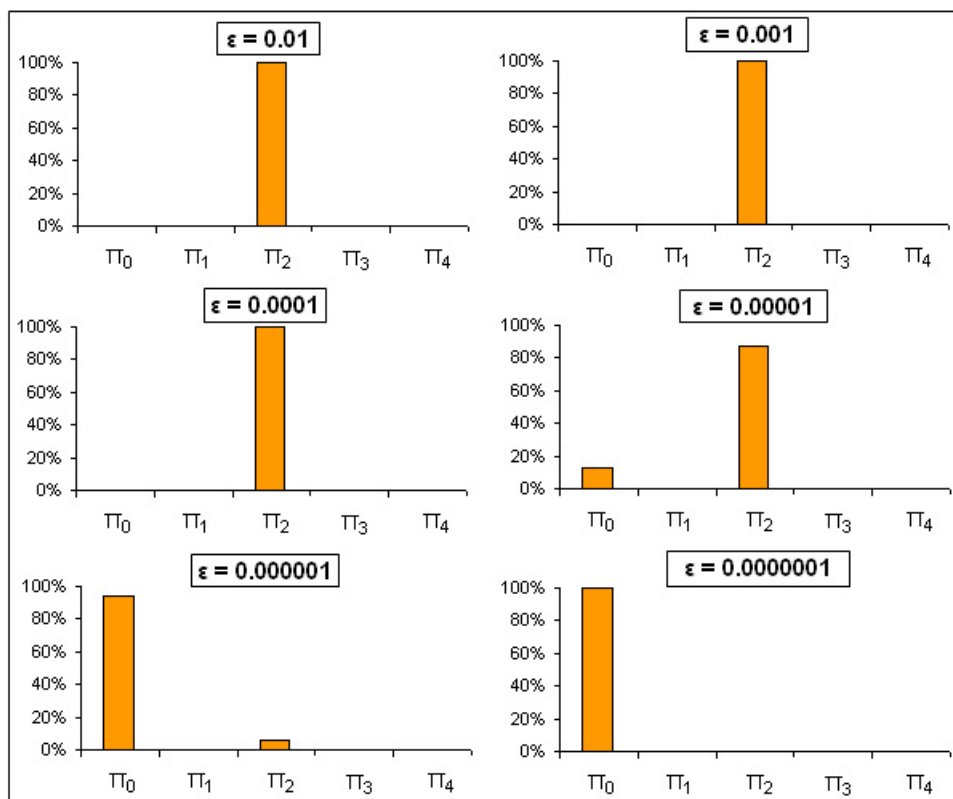
⁵⁴We can also prove that B_0 is the stochastically stable state using proposition one from my previous paper 'Three's a Crowd'. Here $S_1 = 1, S_2 = 2, S_3 = 2$ and $S_4 = 6$ such that all $S_j > 0$ and therefore E_0 is the stochastically stable state.

⁵⁵Let us note that there are several populations where stochastic stability does reflect equilibrium selection under small positive noise levels. The aim of the paper is to show that are *also* many populations where the opposite is true, and for particularly small positive noise values. Section 6 discusses the prevalence of our main result.

4.5 The Limitations of Stochastic Stability

In this section we show even more startling results with populations subject to *extremely* small mutation levels being located in a completely different neighbourhood to the stochastically stable equilibria.

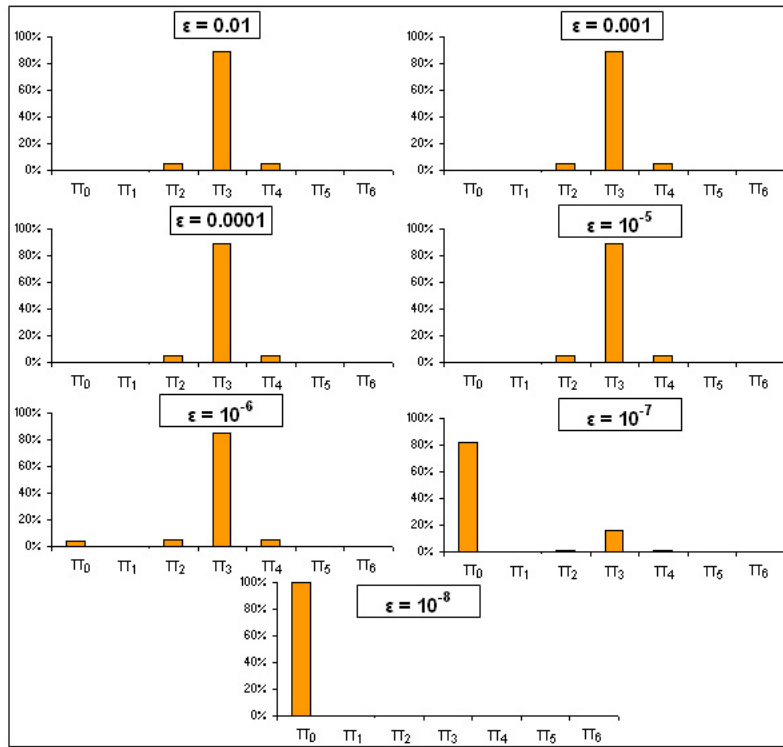
The charts below show the stationary distribution of another four type population with each type equally represented such that $\alpha_i = \frac{1}{4}$ and thresholds $p_1 = 0.2$, $p_2 = 0.31$, $p_3 = 0.69$ and $p_4 = 0.81$



Remarkably the stationary distribution at mutation levels as low as one in ten thousand is completely different to the distribution implied by stochastic stability. As such vanishing noise analysis cannot in any way be seen as an accurate representation of the long run distribution of this

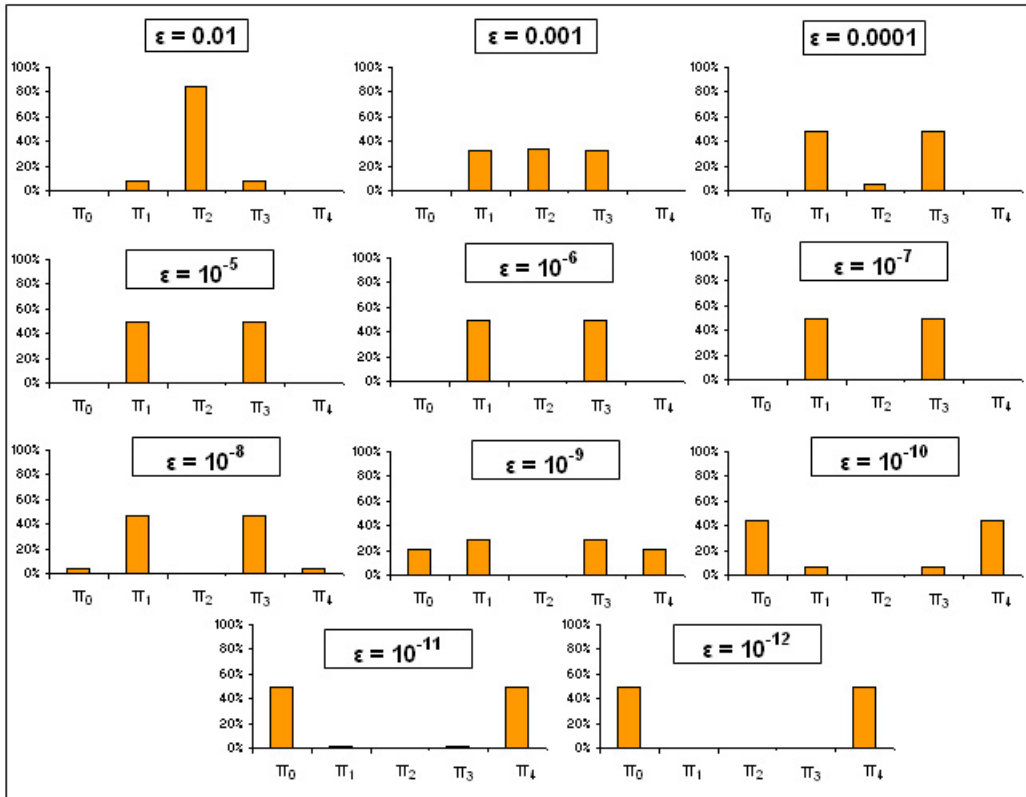
population under even very small mutation rates. Under high noise levels one would naturally expect the process to locate in central equilibria, what is surprising here is that we find remarkably different results for such small levels of ε .

We continue by showing similar results are repeated for a population with six types equally represented in a 102 player population with thresholds $p_1 = 0.12, p_2 = 0.22, p_3 = 0.42, p_4 = 0.58, p_5 = 0.78$ and $p_6 = 0.88$. We can see below that this population exhibits the same characteristics as our first two examples.



We now present one of our most extreme examples below with a 100 player population consisting of four player types equally represented with preferences $p_1 = 0.13, p_2 = 0.38, p_3 = 0.62$ and $p_4 = 0.87$. Amazingly the stochastic stable equilibrium only appears in the stationary distribution

with any positive weight when the mutation rate reaches values as low as 10^{-8} .



The results show conducting stochastic stability alone can be a very dangerous method of selecting the long run location of certain models even with very small levels of noise.

4.6 The Prevalence of Stochastic Stability Misrepresentation

We now look to assess the frequency of populations with small noise levels who are misrepresented by stochastic stability analysis. To do this we simulate many different populations which have the same number of player types but have players with different preference values. We then determine the percentage of populations where small ε results are in fact different to stochastic stability results.

Each population consists of 100 players and the player types are equally represented such that $\alpha_i = \frac{100}{m+1}$.⁵⁶ We wish to assess populations with a variety of different preference values. As such each type's threshold's are drawn from a uniform distribution on the interval given by our constraint $p_k < \sum_{i=1}^{i=k} \alpha_i < p_{k+1}$. And so for each population p_1 is drawn from $U[0, \alpha_1]$, and generally p_i is drawn from $U[\alpha_{i-1}, \alpha_i]$.

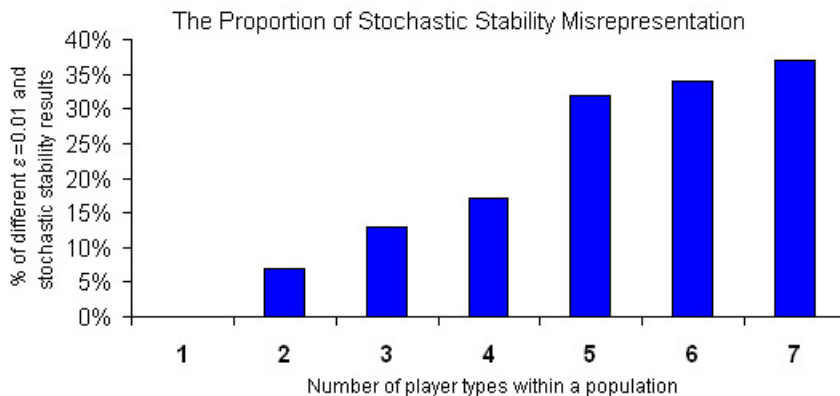
Each set of p_i values represents a single population, and we calculate the stationary distribution of each population for $\varepsilon = 0.01$, to assess small positive noise levels, and for $\varepsilon = 10^{-11}$ to determine the stochastically stable equilibrium.⁵⁷ We then assess whether the neighbourhood with the highest stationary weighting, π_i , is the same for both values of ε or not. We produce these calculations for 10000 populations and determine the percentage of populations which have different $\varepsilon = 0.01$ and stochastic stability results.

We now present our results of the percentage of populations with

⁵⁶For $m = 2$ and $m = 5$ we have $N = 102$, and for $m=7$ we have $N = 98$ as $\alpha_i N$ must be an integer. We have chosen not to vary N and $\alpha_i N$ here and leave such investigations for future research.

⁵⁷This is generally a small enough value of ε to find the stochastically stable equilibrium.

different $\varepsilon = 0.01$ and stochastic stability results, for increasing levels of heterogeneity.



The one type population corresponds to KMR’s original homogenous model and here stochastic stability results do match those with positive noise levels as there are no polymorphic states. However when a degree of heterogeneity is introduced into the model we can see that stochastic stability is a poor representation of *many* populations subject to small noise levels. As such the results of our previous section are not a freak occurrence. We find 17% of four player type populations have different $\varepsilon = 0.01$ and stochastic stability results. As such our results from section 4 and 5 are not a fluke. Indeed for higher levels of heterogeneity we find more startling results, over a third of six player type populations have different positive noise and stochastic stability results. This is an extremely high occurrence of stochastic stability misrepresentation, and further demonstrates the weakness of vanishing noise analysis.

We now discuss why we observe different stochastic stability and positive noise level equilibrium selection results, and why this difference occurs more readily in populations with higher levels of heterogeneity.

4.7 Discussion of the Results and Dynamics

In this section we wish to address the reasons for the difference between stochastic stability and small noise level results, the direction of the change and why such small mutation rates can yield such a significant disagreement. We also discuss why populations with higher levels of heterogeneity are more likely to be misrepresented by stochastic stability.

The stark difference in results is due to the force of using stochastic stability as a limit. Allowing noise to completely vanish overwhelms two effects that are present and significant under positive noise even when the levels of that noise are extremely small. The first effect is the smaller of the two and comes from the fact that simultaneous opposing mutations can cancel each other out. This effect is overwhelmed by stochastic stability, and when the effect is present under positive noise it favours states of co-existence.

The main positive noise effect however comes from the combinational dynamics of the process. The number of potential mutations in order to complete a basin jump varies between different equilibria, which has a large influence on transition probabilities under *any* positive noise levels, yet is completely overlooked when taking the limit of noise to zero. Under positive noise a mixture of combinational forces and the size of basin jumps will decide equilibrium selection, however under stochastic stability only the size of the basin jumps is of importance.

For instance when the process is in B_0 all N agents can jump towards B_1 , whereas at B_1 there are only $\alpha_1 N$ agents who can jump towards B_0 . As such due to the larger combinational forces in B_0 the probability of x mutations is strictly greater than x type 1 mutations at B_1 towards B_0 for any noise level. However as the the noise level vanishes the combinational

forces are overwhelmed and probabilities converge to the same value.

The combination effect is always in favour of co-existence states under positive noise levels and as such our results cannot go the other way. If stochastic stability does select a state of co-existence then the long run location of the process under positive noise cannot be a monomorphic neighbourhood.

For populations where the B_0 basin escape requires slightly more mutations than the jump from B_1 to B_0 , due to the combination effect for some positive noise levels the process will spend more time in B_1 than B_0 . Stochastic stability instead tells us more time will be spent at B_0 than B_1 as the basin B_0 basin jump is larger. As such there is a range of type one thresholds p_1 between the equilibria E_0 and E_1 where some positive level of noise will yield probabilities $p_{01} > p_{10}$ while stochastic stability tells us $\lim_{\varepsilon \rightarrow 0} \frac{p_{01}}{p_{10}} \rightarrow 0$, and as such the stationary distributions can be very different. I shall denote the range as ε_0^+ .

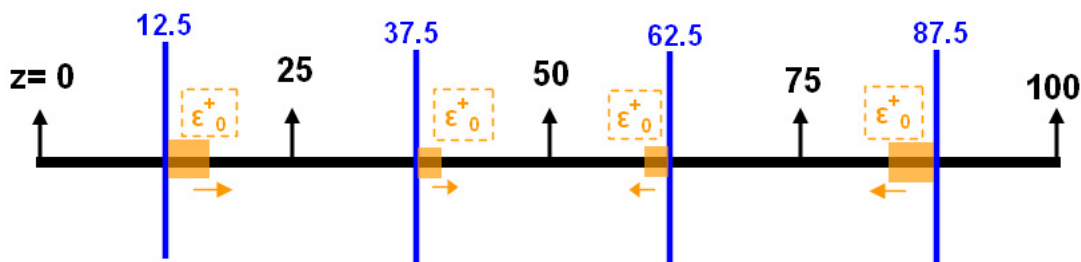
It is the heterogeneity of our model which is the key to why we obtain our results, even at very small mutation levels, for two reasons both associated with combinational forces.

Firstly there exists an ε_0^+ range between every two neighboring equilibria.⁵⁸ In a heterogenous population there are many equilibria and therefore many ε_0^+ ranges. If one threshold is within an ε_0^+ range then this in itself can lead different stochastic stability and positive noise results. If more than one threshold is within an ε_0^+ range then the effect can be extenuated, and stochastic stability and positive noise results can be very different for very small values of ε . In each case the effect of a threshold

⁵⁸The only exception being when two neighbouring equilibria are situated equally far from $p_i = 0.5$. Here the number of possible mutations from either equilibria are identical, and so there is no combination effect present here.

being in the ε_0^+ range is that more weight is given to more central neighbourhood than stochastic stability would suggest.

The state space with ε_0^+ ranges for the 100 player population with 4 types equally represented is illustrated below.⁵⁹



Secondly the linked basins of a heterogenous population means that an equilibria's long term weighting is dependant upon the interaction of transition probabilities across the whole state space, not just the equilibria's own basin jump probabilities. For instance B_0 's long term weight in the stationary distribution is in part determined by E_2 's basin jump probabilities as essentially $\pi_0 = \frac{p_{10}}{p_{01}} \frac{p_{21}}{p_{12}} \pi_2$.⁶⁰ If the jump $B_1 \rightarrow B_2$ requires a relatively small amount of mutations⁶¹ but slightly more than $B_1 \rightarrow B_0$ then under some positive noise levels due to combinational forces $\frac{p_{21}}{p_{12}} \frac{p_{10}}{p_{01}} > 1 \Rightarrow \pi_0 < \pi_2$. However because the basin jump $B_1 \rightarrow B_2$ requires slightly more mutations the results, for some very small ε the results must switch such that $\frac{p_{10}}{p_{01}} \frac{p_{21}}{p_{12}} < 1 \Rightarrow \pi_0 > \pi_2$. The many basins of a heterogenous population create various scenarios in which such switches

⁵⁹The ε_0^+ ranges are drawn wider near the monomorphic equilibria than the middle equilibria. To see why consider that the number of possible mutations from E_0 to E_1 is 100 and the number of possible mutations from E_1 to E_0 is just 25, a difference of 75. However the number of possible mutations from E_1 to E_2 is 75 and the number of possible mutations from E_2 to E_1 is 50, a difference of 25. As such the ε_0^+ range is larger between E_0 and E_1 than E_1 and E_2 .

⁶⁰Cross basin jumps, p_{20} etc, being regarded as negligible here.

⁶¹Such that $\frac{p_{10}}{p_{01}} > 1$ and $\frac{p_{21}}{p_{12}} > 1$.

occur at *extremely low* ε rates as shown in the calculation examples.

The linked basin effect and the ε_0^+ range effect *both* increase with the level of heterogeneity of a population, which explains why we find a greater proportion of populations are misrepresented by stochastic stability when the populations have more player types. For higher levels of heterogeneity we have more basins of attraction, and therefore a greater scope for the linked basin effect to produce different stochastic stability and small noise results. With higher levels of heterogeneity there are also more ε_0^+ ranges in the state space and therefore more populations are affected by player thresholds landing within an ε_0^+ range. The combination of both effects produces our finding that 56% of six player type populations are misrepresented by stochastic stability.

And so both when thresholds fall in ε_0^+ ranges and when linked transition probability interactions are sensitive to ε rates, stochastic stability can be a very poor representation of many populations subject to even very small levels of noise.

4.8 Conclusion

We have shown through calculations that stochastic stability can be a very dangerous method of representing a heterogenous populations subject to small levels of noise. Most surprisingly we have found that even with populations subject to *extremely* small noise levels stochastic stability can completely misrepresent the long run location of the population. We also find that such a representation is not rare and occurs in over a third of populations. We also find that when there is a discrepancy between the two methods it is the states of co-existence that stochastic stability overlooks.

And so this paper wishes to alert the research community of the dangers of conducting vanishing noise analysis alone, without consideration of how results may change under a rather small amount of positive noise.

References

- Anwar, A. (1999) "On the Coexistence of Conventions (2002)". Journal of Economic Theory.
- Alan Beggs, (2002). "Large Deviations and Equilibrium Selection in Large Populations,". Economics Series Working Papers 129, University of Oxford, Department of Economics.
- Blume, L. (1999): "How Noise Matters" Working Paper.
- Bergin, J., and B. Lipman (1996): "Evolution with State-Dependent Mutations", *Econometrica*, 64, 943-956.
- Ellison, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047-71.
- (2000):"Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution" *Review of Economic Studies*, 67,

17-45.

Granovetter, Mark. (1978). "Threshold Models of Collective Behavior." *American Journal of Sociology*, 83: 1420-1443.

Hojman D, (2004) "Interaction Structure and the Evolution of Conventions."

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

Kolstad, I. (2003): "Viscosity and dispersion in an evolutionary model of learning."

Myatt, D.P and C. Wallace (1998): "Adaptive Play by Idiosyncratic Agents," *Discussion Paper, Nuffield College, Oxford*.

Norman, T. (2003a): "The Evolution of Co-ordination under Inertia."

Robson, Arthur J. and Vega-Redondo, Fernando, (1996): "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *Journal of Economic Theory*, Elsevier, vol. 70(1), pages 65-92.

Samuelson, l., (1994):"Stochastic Stability in Games with Alternative Best Replies," *Journal of Economic Theory* 64 , 35-65.

Sandholm, W. H. (2006). "Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium". Unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2007). "Population games and evolutionary dynamics". Unpublished manuscript, University of Wisconsin.

Schelling, Thomas C. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1: 143-186.

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

(1990) "Stochastic Evolutionary Game Dynamics" with Dean P. Foster.

Theoretical Population Biology, 38, 219-232.

5 Models of Evolution and Social Thresholds

Abstract

This literature review looks at a series of papers whose models are driven by social thresholds, the idea that people's decisions are strongly influenced by the actions of society as a whole. The literature began with Schelling's investigative paper on social integration proceeded by a formal model by Granovetter on group action. Later papers delve into the influence of social networks and the speed of evolution. We give a thorough description of each key paper and provide insight into the assumptions that drive the results of the varying models and the overall literature.

5.1 An Introduction to Social Threshold Models

In his ground breaking paper 'Dynamic Models of Segregation' Thomas Schelling first looked at a game which modeled a society consisting of white and black people with the simple requirement that in a certain neighbourhood, individuals did not want to be in the minority. When an individual considered a neighborhood they looked at the proportion of their own colour and if this fell below a certain threshold they would not join the neighborhood. Schelling went onto demonstrate this simple and seemingly insignificant preference can lead society to extreme segregation, and his work began a literature of models based upon social thresholds. Social threshold models are driven by the assumption that people are strongly influenced by other members of society, and when enough people take up an action others will may also be persuaded to join in. Mark Granovetter's first formal model of social thresholds, 'Threshold Models of Collective Behavior', explains the idea extremely well. If one person starts a riot then another person may join in. If 50 people have joined the riot already this may persuade another 10 to riot, and eventually through this dynamic an entire group may join in. Grovetter's paper gives us a first insight into the importance of the distribution of thresholds in the population, and this is the second paper we study.

One can naturally ask whether people are in fact strongly influenced by the actions of others and indeed whether threshold models can be an accurate representation of individual choices in society and group outcomes. It can be in little doubt that in a variety of situations people are influenced by the actions of others. For instance in some cases the fact that a large proportion of people are taking up a new innovation may imply they know something you don't, and as such persuade you to also buy

the innovation. Indeed a disease will spread at a faster rate the more a population is infected. Also one may only want to speak out on an issue, or vote on an issue if they believe other people hold the same view. Obviously there are many examples where individuals don't care what others are doing, however there are certainly many interesting cases where people do care about the actions of others, certainly enough to justify the focused research on social threshold models we survey in this review.

The previous chapters of the thesis are studies of populations of players who have a range of thresholds and this is the motivation for the particular literature review of social threshold models chosen here. However one criticism of the literature on threshold models is that people change strategies just because other people have changed without any economic justification for why this should happen. Players often do not have utility functions such that their chosen strategies cannot be judged to be optimal or not, and as such it can be argued that economists should not be concerned with pure threshold models.

However the threshold nature of the models of our previous chapters are not assumed but produced from populations who play a co-ordination game with actual payoff matrices and player utility functions. Indeed the general level of heterogeneity of the populations results in a distribution of thresholds of the type assumed by Granovetter initially and others later. Therefore, as economists we are very interested in the pure threshold models studied in this review.

One application of interest which demonstrates the link of the previous chapters to the current review is the mobile phone market. In many markets calls to your own network are significantly cheaper than calls to other networks, creating a repeated co-ordination game. The more peo-

ple in a population on your network the higher your payoff will be, such that it may be optimal for you to switch networks if many other people switch strategies, leading to player thresholds. Furthermore in our previous chapters we also allow players to have personal preferences for a particular network, and as such different players will have different optimal actions based upon the play of the population as a whole, generating a range of player thresholds of the type assumed in many of the papers studied in this review. The only significant difference between the models in my previous chapters and those of the current literature review is that in the current literature agents can adopt a strategy but often cannot change their mind, whereas in my model it is possible for agents to switch back and forth between two strategies.

Returning to the contents of the review the third paper we look at is Micheal Macy's paper 'Chains of Co-operative Action: Threshold Effects in Collective Action'. In Granovetter's paper he often finds populations which get stuck at very low adoption levels, or no adoption at all and it is this issue which Macy looks to address. By introducing a threshold distribution that can evolve over time Macy's model can create a mechanism for a population to escape from low levels of adoption. We go onto examine Mark Gronovetter's second paper co-written with Roland Song, 'Threshold Models of Diversity: Chinese Restaurants, Residential Segregation and the Spiral of Silence'. This paper attempts to create a formal mathematical account of Schelling's initial ground breaking paper on social segregation, recreating many of his simulated results.

Next we consider Peter Dodds and Duncan Watts paper, 'Universal Behavior in a Generalized Model of Contagion' where agents can remember a number of previous interactions and their threshold can therefore be

breached via a series of contacts with other members of a population. In a quite complicated structure the authors are able to determine which of three classes a model's equilibria will fall in just by looking at two parameters. Here it is those most easily influenced in the population who are key to the diffusion dynamics.

The penultimate paper we study is the social threshold section of Peyton Young's recent paper 'Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence and Social Learning'. Peyton's paper is original in that he focuses on the speed of adoption. He demonstrates that the speed of adoption is dependant upon the initial conditions of a society and shows if the early adoption levels are high enough the process will sustain a period of super-exponential growth.

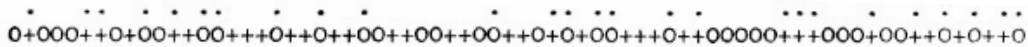
The last paper we study is Thomas Valente's 'Social network thresholds in the diffusion of innovations'. His paper is unique as he conducts an empirical investigation on the diffusion of innovations with data from three separate cases. The paper assesses the influence of social networks on the adoption of new ideas in a society. He finds amongst other results that those individuals who have the most affect upon the adoption of innovations in a population are generally both conservative and consistent in their views and actions.

The first paper we begin with is Schelling's corner stone paper, and the rest of the literature review continues in the order described by the introduction.

5.2 Dynamic Models of Segregation

— Thomas.C.Schelling 1971

In this paper Schelling pioneers the idea of social thresholds in dynamic models of individual choice. The thresholds here derive from an agent requiring a certain number of their own type to be in their neighbourhood in order to be content. Schelling considers two groups of players who decide which neighbourhood they wish to live in. He begins with an elegant one dimensional model consisting of a line of players of both groups. The diagram below demonstrates the model, with stars and crosses representing the two separate groups in a random order, and is taken from Schelling's paper.



An agent's neighbourhood is defined initially as an agent's immediate four players to their left, four players to their right and themselves.

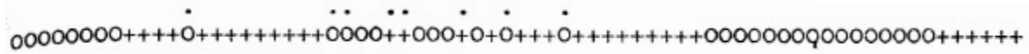
The key feature of the model is that in terms of their neighbourhood individuals do not want to be in the minority.⁶² Therefore a star requires at least four of his eight neighbours to be stars in order to be content, as he also includes himself. This introduces the *social threshold nature* of the model which inspired a whole literature.

If an agent is not content in their current neighbourhood then they can move. Those who are unhappy in the first diagram have been denoted with a dot above them. The process moves from left to right in that the first discontented agent to move is the cross who starts out second in from the left in the diagram. A discontented agent must move to the closest

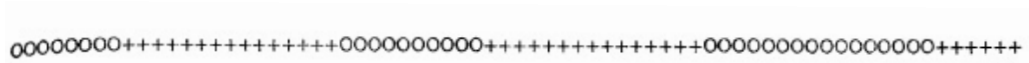
⁶²But agents do not necessarily have a preference to be part of a strong majority.

neighbourhood that satisfies his preferences not to be in the minority. When an agent moves this may cause some previously content agents to become discontent, a key rule is that such an agent can move but only after the all the previously discontented agents move (all those with a dot above them in the initial diagram). Similarly some previously discontent agents can become content when an agent moves, if this is the case the previously discontented agent will not move.

Following these rules, once those with the dots in the initial diagram have become content we obtain the new social distribution below.



There are still eight discontented agents due to the other agents' moves. Once these eight move we reach a state of equilibrium in which all agents are content which is shown in the diagram below.



The startling and fascinating result Schelling obtains is the overwhelming segregation of the two groups, when the *only requirement* agents have is not to be in the minority. Agents would be perfectly content in a population with stars and crosses alternating, yet when left to the dynamics of the model extreme segregation results. Schelling verified his key result with many simulations to show the initial distribution in the first diagram does not drive the main result.

The model lends itself to many variations, Schelling considers neighbourhood size, neighbourhood preferences, the ratio of the two groups, the rules of movement and the initial distribution. Schelling finds that the

main result is robust to most assumptions on these variations. An interesting result is obtained from restricting movement by putting a limit on an individual's travel, when there exists a minority group. Surprisingly agents achieve their desired neighbourhood faster and without travelling as far as they would in a free movement model. The restriction channels society into small, more frequently occurring clusters.

Schelling develops his main result in the rest of the paper by looking at a two dimensional form of the model, consisting of agents living on a checker board type arrangement. Schelling repeats his main result that just a small individual preference not to be in the minority leads to significant segregation holds in the two dimensional model. Schelling continues to investigate several variations on this model such as intensity of the population densities and individual demands for social integration with interesting results.⁶³

In this paper Schelling introduces the concept of social threshold levels when an agent looks at the proportion of two groups in a neighbourhood. Schelling's concept of social thresholds could then be applied to any model in which agents care about the proportion of agents already taking an action. Granovetter's next paper details how such a concept could be mathematically represented, and applies his model to riots amongst other interesting examples.

⁶³In the final section of the paper Schelling changes the definition of neighbourhood from those located next to an agent to a set area, defined as a bounded neighbourhood model, and investigates a series of topics with varying results.

5.3 Threshold Models of Collective Behavior

— M.Granovetter 1978

Granovetter is one of the pioneers of theoretical models which are driven by individual threshold levels, where an individual will only take an action when a certain number of other people have already taken the action. He looks to continue from Schelling's original paper by forming a mathematical model which holds the essence of Schelling's paper with social thresholds at the centre of the dynamics.

Granovetter introduces the paper with the example of a riot. Granovetter asks the question with 100 people in a square, with at least one person willing to begin the riot, will others join in next and more after them in a domino type effect, and how many of the 100 will end up taking part in the riot. Importantly people will join the riot only once a certain number have already joined, this level being the social threshold for each individual. The key to the model is that people have different threshold levels, and it is the nature of this heterogeneity that determines the resulting size of the riot. The model can be extended to many different and interesting cases such as the take up of innovations, rumours, strikes, voting and migration.

Granovetter considers a clear and simple case of a uniform distribution of thresholds such that one person wishes to begin a riot having a threshold of 0, and another person has a threshold of 1 person in order to join the riot, another with threshold of 2 and so on with the most resistant joining in only if 99 people have joined in already. In this case it is clear to see due to a domino effect that all 100 people will join the riot in the end.

However Granovetter then considers that the person with threshold 1 is instead replaced with a person of threshold 2. We now have a completely

different result where we have a riot of just one person with no-one else willing to join in. Granovetter uses this simple case to illustrate that drawing conclusions of individual views from group outcomes can be very misleading indeed.

Formally Granovetter gives each individual in a population a threshold level x , such that an individual will only take an action when at least x players have already taken action x . $F(x)$ is the cumulative distribution frequency and $f(x)$ is the frequency density function. The proportion who have adopted the action at time t is denoted $r(t)$.

Consider that one person has joined the riot in period one, and there are ten people with threshold 1. Then next period all these 10 people will join to make 11 in total. In the following period all those with thresholds less than 11 will join in. And so if at any time t $r(t)$ agents have joined, then the amount of people who will have adopted in the next period is given by the difference equation:

$$r(t + 1) = F(r(t))$$

And so the process will come to rest at equilibrium when it reaches the first fixed point of $F(r)$, where $F(r) = r$.

Granovetter decides to focus on thresholds that are normally distributed in the population of 100, leaving other distributions to further research. An example of a normally distributed threshold density with mean 25 is used, and Granovetter considers the resulting equilibrium level R_e for varying levels of standard deviation levels, σ . The graph below is taken from the paper and displays Granovetter's main result.

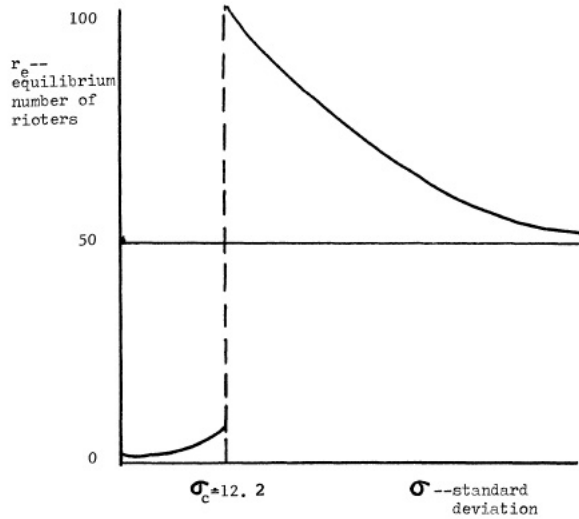


FIG. 2.—Equilibrium number of rioters plotted against standard deviation of normal distributions of thresholds with mean = 25, $N = 100$.

For low σ , there is little mass in the left tail of the threshold distribution, and so the process stops early on. For a critical level of $\sigma \simeq 12$ and beyond there is enough mass in the left tail such that process will create a knock on effect such that at least half of the population will adopt. For σ just above 12 adoption will be close to 100%. As σ increases, the decreasing slope of the right hand side of the normal distribution causes the process to stop at a decreasing equilibrium level of adopters.

Granovetter’s main point is that a very small change to the individual views of the population can lead to radically different outcomes, and so inferring individual views from group outcomes is a dangerous business. Next Granovetter looks at the introduction of social networks into the population, so that an individual will put different weights on different people joining the riot. For instance Granovetter considers a population who count friends joining the riot twice as much as strangers joining. There are obviously any number of social networks possible. Granovetter also introduces perturbations into the model, allowing the threshold distribution to be subject to small shocks. Characterising stable equilibria mathematically is difficult and most of Granovetter’s results are produced using computer simulations.

Granovetter's general results show that the nature of social networks can have a highly significant effect upon the resulting equilibrium, and interestingly states that the modal equilibrium result is one player joining the riot, under a variety of network and perturbation combinations. The stochastic nature of the threshold distribution means that the probability of no agents with threshold 0, or some with 0 but none with threshold 1 is in fact often greater than 50%, explaining the low model equilibrium result. And so with perturbations collective action may be unlikely to occur even if most of the population would happily join in once the process gets going. Therefore we can see that the results from the initial threshold model, especially for cases resulting in high adoption levels are unlikely to reflect a population subject to thresholds that may evolve over time. In the remainder of the paper Granovetter discusses a variety of views many of which we have in part discussed in the introduction of this paper. In this paper Gronovetter introduces a rather elegant mathematical representation of a population of individuals who are influenced by what others are doing and as such have social thresholds. Gronovetter produces some interesting results on the nature of the resulting equilibrium and builds upon Shelling's original ideas in a more mathematical way.

5.4 Chains of Co-operative Action: Threshold Effects in Collective Action - M.W.Macy 1991

This paper follows on directly from Granovetter's 1978 paper. Primarily Macy looks to address the issue of low uptake cases, where the population is stuck in a non-cooperative equilibrium. In order to address this issue Macy's main change to Granovetter's model is to introduce a stochastic rather than deterministic distribution of thresholds. As such agents will adjust their threshold levels in a direction dependent upon the payoff of the last two periods. This allows agents to gradually reduce their thresholds when stuck in a low uptake case, until the threshold distribution evolves such that there is now a critical mass willing to adopt which catapults the population into a state of increasing adoption levels. Essentially Macy's stochastic threshold distribution can fill the gaps in Granovetter's static threshold model allowing a population to escape non co-operative equilibrium.

As well as introducing an evolving threshold distribution Macy also adds a variety of other factors in his model, and considers a formal public goods game with individual contributions, a production function and a method of distributing the public good to individuals. All of the results in the paper are from simulation alone. We next look at the specifics of the model.

Firstly rather than agents having a threshold at which they will definitely contribute to a public good, agent j has a probability P_j of contributing at any point depending upon their threshold T_j , the participation rate π and a slope parameter M .

The specific logistic functional form is given below:

$$P_j = \frac{1}{1 + e^{(T_j - \pi)M}}$$

such that an agent has a 0.5 chance of adoption when the participation rate reaches T_j .

Different agents contribute different amounts C_j according to

$$C_j = R_j N V_i$$

where R_j is agent j 's share of total resources N and $V_j = \{0, 1\}$ represents whether the agent has contributed or not.

And agent receives a share of the public good, S_j , given by

$$S_j = \frac{L N I_j}{N(1 - J)} - C_j$$

where L is the production level, I_j is an individual's level of interest in the public good and J represents a jointness of supply, such that higher values of J will reap higher co-operation rewards.

The production level L is given by the production function below,

$$L = \frac{1}{1 + e^{(0.5 - \pi)10}} - \frac{1 - X}{2}$$

where X can be varied to make the public good game a public bad game.

We now look at the Macy's learning algorithm which is the main development from Granovetter's 1978 paper. Macy introduces an outcome function given by

$$O_{ij} = \frac{E_j(2S_{ij} - S_{i-1,j})}{3|S_{max}|}$$

and the evolution of an agent j 's threshold is given by

$$T_{i+1,j} = T_{i,j} - V_{i,j}[O_{i,j}(1 - (1 - T_{i,j})^{|0_{i,j}|})] + (1 - V_{i,j})[O_{i,j}(1 - (1 - T_{i,j})^{|0_{i,j}|})]$$

such that an agent's threshold will fall if he co-operates and is rewarded, or if he does not co-operate and is punished.

It is the assumption that agents' thresholds can change and evolve that drives Macy's results and leads to different conclusions from Granovetter's paper. The paper contains two key results.

Macy's main result is that in his model a population can overcome the low uptake levels found in Granovetter's examples. If the population is stuck in low uptake levels over time the threshold distribution will shift to the left until it reaches a critical point which sparks a period of mass co-operation, explaining the 'chain of co-operation' terminology of the title. The initial period of low co-operation can be seen as pulling back an elastic band, with the release of the band being chain reaction of co-operation.

Macy also compares the model when agents move simultaneously and sequentially, termed parallel and sequential choice respectively in the paper. Under simultaneous decisions the population is in isolation and often struggles to escape a non co-operative equilibrium. This is because equilibrium escape requires a number of agents to change their strategies at the same time, and such a random fluctuation is unlikely.

However under sequential moves a single agent's choice to contribute will lower other agent's thresholds through the evolving threshold dynamic, thereby making them more likely to contribute next period, and so here the population can often escape the non co-operative equilibrium.

The two main results are show in the diagram below.

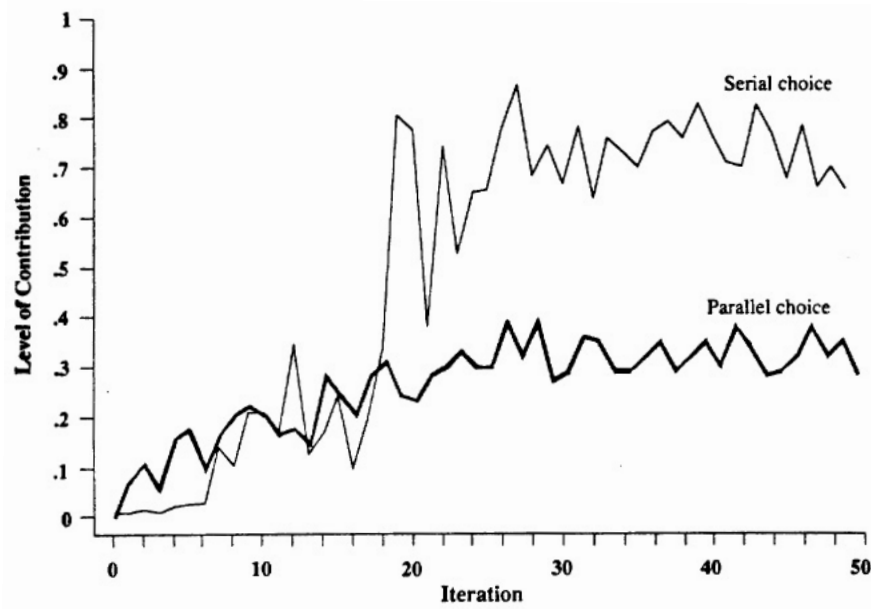


Figure 1. Serial and Parallel Choice in Collective Action ($N = 100, J = .5, X = 0, C = \{0,1\}$)

Macy like Granovetter looks at the effect of the population's social structure in which agents are only influenced by their friends. Macy considers a strong tie configuration in which agents are in small groups, and are only influenced by members of their group, such that agents are always friends with their friend's friends. Macy also considers a weak configuration, in which agents are paired with others in the population, but agents are often not friends with their friend's friends. Interestingly a population with weak social ties escapes low uptake levels much more readily than a population with strong social ties. Indeed the results with strong ties are similar to those under simultaneous moves. Cliquishness means that isolated groups do not influence each other, and so equilibrium escape requires separate groups to change strategies and contribute simultaneously in much the same as with simultaneous decisions.

Macy finishes the paper by showing that high adoption levels are generally robust to low jointness of supply, few positive rewards and divisible contributions. Macy also looks at the effect of correlation between an individual's resource level, R_j , and their interest in the public good, I_j , finding correlation levels do not prevent high adoption level results. Furthermore different R_j and I_j distributions reveal that it is often those who have the most interest in the game, but not the most resources who often bear the burden of contribution.

The main contribution of Macy's paper is to show that allowing a population's threshold distribution to evolve over time allows the gaps of Granovetter's model to be filled which creates a mechanism for a population to escape from low levels of adoption.

5.5 Threshold Models of Diversity: Chinese Restaurants, Residential Segregation and the Spiral of Silence - Granovetter and Song 1998

In this paper Granovetter and Song attempt to formalise mathematically some of Schelling's results. Schelling's original paper detailed an open model and showed a variety of interesting outcomes including the main result that just a small individual preference not to be in the minority can lead to significant social segregation. Granovetter and Song form an exact mathematical model and look to see amongst other things whether Schelling's main result is produced by their model.

Granovetter and Song define two separate groups as white and black people as Schelling often does. N_b and N_w denote the total number of black and white people who could move into a neighbourhood from a population of N players. As such the actual neighbourhood size can vary over time. Each individual has a social threshold p_w dictating the minimum proportion of whites needed in a certain neighbourhood in order for the agent to be content. $F(p_w)$ denotes the fraction of whites with thresholds below or equal to p_w . Furthermore n_b and n_w denote the number of black and white people in a neighbourhood such that $p_w(t) = \frac{n_w(t)}{n_w(t)+n_b(t)}$ details the proportion of whites in a neighbourhood at time t , where time is taken to be discrete.

If a neighbourhood was 25% white at time t and $F_w(0.25) = 0.6$ then 35% of the white population would wish to join the neighbourhood, and therefore generally we have

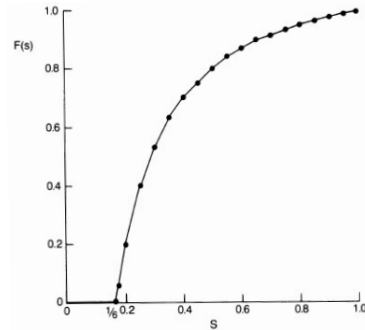
$$\begin{aligned}n_w(t+1) &= F_w[p_w(t)]N_w \\n_b(t+1) &= F_b[p_w(t)]N_b\end{aligned}$$

Equilibrium results when both $n_w(t + 1) = n_w(t)$ and $n_b(t + 1) = n_b(t)$, and w and b represent the number of black and white people in a neighbourhood in equilibrium. The resulting algebra from the equilibrium conditions is complex and so for clarity of explanation we concentrate on Granovetter and Song's graphical demonstration of their model and results.

Granovetter and Song use curves illustrating the maximum number of group A a certain group number of group B can withstand in a neighbourhood, these type of curves were first being introduced by Schelling. The maximum number of white people a group of black people in a neighbourhood can tolerate is given by the black person with the highest threshold. Given the neighbourhood is in equilibrium and considering $F(s)$, then s is the highest threshold among a proportion k of the population where $F(s) = k$. We have it that in equilibrium $k = \frac{b}{N_b}$ and $s = \frac{b}{w+b}$.

It is then possible, given b , N_b and F to deduce the highest threshold proportion and the maximum number of white people a certain group of black people can tolerate. If $F(0.4) = 70$, then 70 black people out of a population of 100 can live in a 40% white neighbourhood. Moreover, solving $\frac{70}{70+w} \leq 0.4$ gives us that $w \leq 105$ is required for 70 black people to be content. Therefore for each value of b , 70 in the example here, the maximum number of white people that can be tolerated is easily computed. Granovetter and Song denote a function $e(b)$, such that $e(x) = y$ details that a maximum number of x black people can tolerate y white people in their neighbourhood. In our current example $e(70) = 105$.

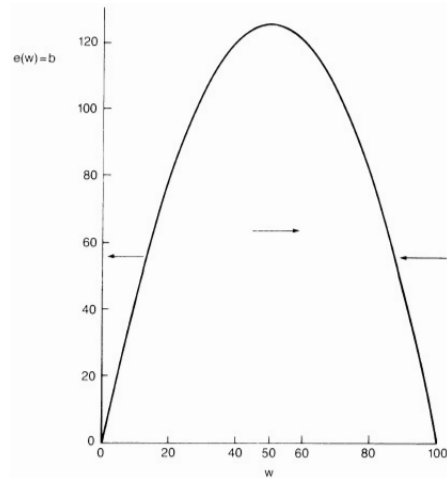
Granovetter and Song consider the cdf threshold distribution used by Schelling, which is given by $F(s) = \frac{1}{R}[1 + R - \frac{1}{s}]$ and a specific example is shown below for $R = 5$.



By solving $\frac{w}{N_w} = F_w[\frac{w}{w+b}]$ for b we can find $e(w)$, and by substituting in the threshold distribution we obtain $e(w)$ as follows.

$$\begin{aligned} \frac{w}{N_w} &= \frac{1}{R}\left[1 + R - \frac{1}{\frac{w}{w+b}}\right] \\ \Rightarrow e(w) &= b = R\left[1 - \frac{w}{N_w}\right]w \end{aligned}$$

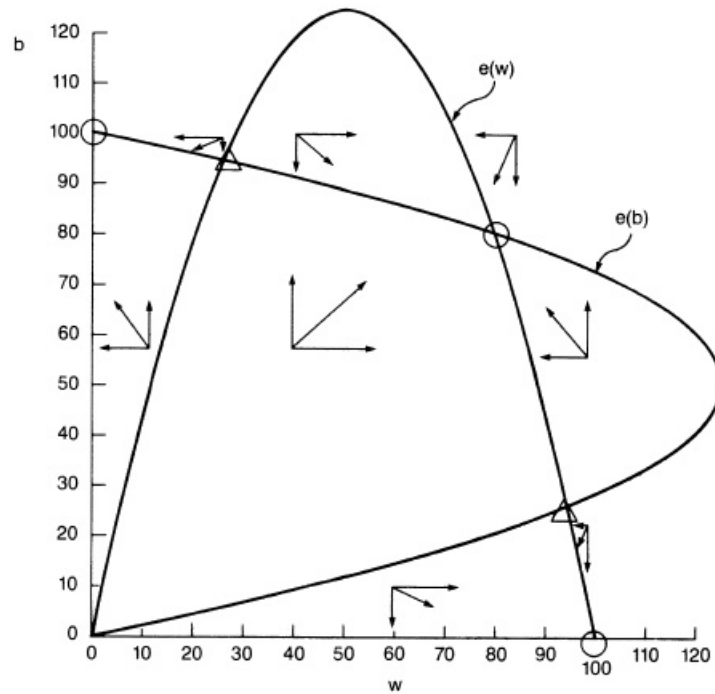
The resulting parabola is graphed below.



The diagram illustrates the movement of white people when in disequilibrium. Within the curve more white people can tolerate a ratio of $\frac{b}{w}$ than exists in the neighbourhood and so more white people will enter the neighbourhood. On the left hand side of the curve the $\frac{b}{w}$ ratio is too low and white people will leave the neighbourhood.

In order to assess equilibria for population as a whole Gravonetter and Song put both $e(w)$ and $e(b)$ curves together, which is represented by the diagram below:

FIGURE 4. $e(w)$ and $e(b)$ for $R = 5$, $N_w = N_b = 100$: Equilibrium curves for whites and blacks with identical straight-line tolerance distributions. Horizontal arrows indicate direction of change in numbers of whites, vertical arrows the direction for blacks, and diagonal (resultant) arrows the direction of overall change. Stable equilibria are circled, unstable ones are enclosed in triangles.



where $(0,100)$ and $(100,0)$ are stable monomorphic equilibria. The intersection of the curves details the three co-existence equilibria, of which only $(80,80)$ is stable. Granovetter and Song do not subject their model to per-

turbations and therefore cannot select between the three stable equilibria in the example shown. However their analysis has succeeded in producing an exact mathematical account for Schelling's paper. Furthermore they do demonstrate in a formal model that neighbourhoods consisting of only one colour are stable equilibria under Schelling's threshold distribution, concurring with Schelling's main result.

We now look at a model in which agents can remember a number of interactions and their threshold can therefore be breached via a series of contacts with other members of a population.

5.6 Universal Behavior in a Generalized Model of Contagion - Dodds and Watts 2004

Dodds and Watts (DW) consider a population of N agents who can be in three mutually exclusive states, S (susceptible), I (infected) or R (removed). DW study a pairwise matching process where an agent i will randomly come into contact with one other agent j each period. If agent i is susceptible and agent j is infected then there is a probability p that agent i will receive a dose, and that dose $d_i(t)$ is drawn from a distribution $f(d)$.

A strong and original feature of this particular threshold model is that an individual can remember doses from the last T periods recalling a total dose $D_i(t) = \sum_{t'=t-T+1}^t d_i(t')$. The threshold nature of the model is given by an individual's *dose threshold* d_i^* which is drawn randomly at the start such that susceptible agents will become infected when $D_i(t) \geq d_i^*$.

Then the probability that a susceptible agent who meets K infected individuals in T periods will become infected is given by

$$P_{\text{inf}}(K) = \sum_{k=1}^K K \binom{K}{k} p^k (1-p)^{K-k} P_k$$

where

$$P_k = \int_0^\infty dd^* g(d^*) P\left(\sum_{i=1}^k kd_i \geq d^*\right)$$

is the average proportion of infected individuals after they receive k doses in T time steps.

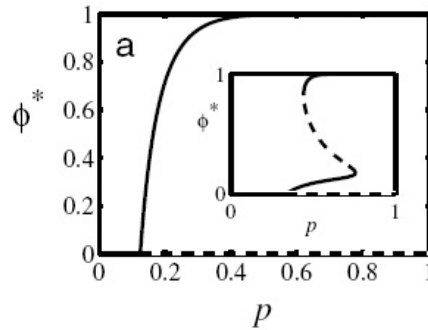
The resulting model is very general but also quite complex, as such DW obtain their results through numerical methods and simulations. DW primarily look at a particular case where an agent will definitely recover if $D_i(t)$ falls back below d^* and recovered agents are susceptible with probability 1. From this model type DW are able to produce an equation

for the steady state proportion of infected agents in the population, which is given below

$$\phi^* = \sum_{k=1}^T \binom{T}{k} (p\phi^*)^k (1 - p\phi^*)^{T-k} P_k$$

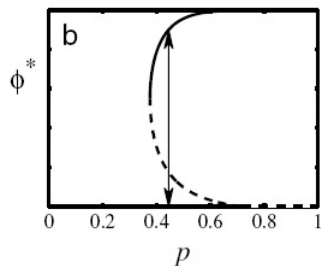
Using this equation DW produce some interesting results. Firstly they show the equilibrium behaviour of their general model falls into three distinct classes. Secondly and surprisingly DW show that the particular equilibrium class depends on just two variables, P_1 , the probability of infection from a single exposure to an infected agent, and P_2 , the probability of infection from two exposures.

The first class is called the epidemic threshold class and an equilibrium will fall into this class if $P_1 \geq \frac{P_2}{2}$, here one encounter with an infected agent is more dangerous than in the other two classes. A graph of the steady state proportion of infected agents for this class is shown below.



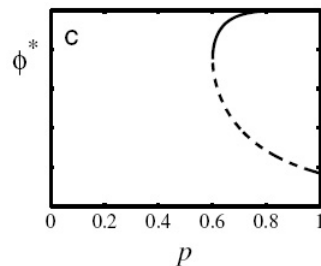
If the probability of receiving a dose, p , is below a critical point p_c the stable equilibrium will be no infections. However if $p \geq p_c$ an epidemic of positive size will occur. The higher p the larger the epidemic will be and for p greater than 0.4 the entire population will be infected.

Equilibria from certain models can fall into a second class type of class if $\frac{P_2}{2} > P_1 \geq \frac{1}{T}$ with the resulting infected population graph is displayed below



For p beyond p_c there exists both unstable (dashed line) and stable (solid) equilibria. The initial $\phi(0)$ level is critical as if the initial $\phi(0)$ falls below the unstable equilibria the infected population will fall to 0. However if the initial $\phi(0)$ is above the unstable equilibria a significant proportion of the population will become infected. The size of the critical mass required decreases as p increases and as such DW term this second class the vanishing critical mass class of equilibria.

Equilibria can fall into a third class when $\frac{1}{T} > P_1$, and the corresponding infected population graph is displayed below.



A large initial $\phi(0)$ level is required in order for the initial seed not to die out and infection to prevail in the long term. As a large initial seed is

required for all values of p and as such DW term this third class the pure critical mass class as opposed to the vanishing second class.

And so DW show that in a quite complicated structure they can determine which of three classes a model's equilibria will fall in just by looking at two parameters, and they continue by showing that these results are robust to varying $f(d)$ and $g(d^*)$ distributions. Their results demonstrate that not all contagion models fall into the same class. As DW state there is little empirical evidence for P_1 and P_2 values. The current literature has said little on the subject of whether exposures are independent events. DW show here for models where $P_1 < \frac{P_2}{2}$ the nature of the interdependencies can have a dramatic effect. In much of the literature persuading the leaders in society is the key to dispersion. Interestingly in contrast here, DW show as P_1 and P_2 are so significant for a model that those most easily influenced could in fact be the key to the diffusion dynamics.

5.7 Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning - P.Young 2008

We begin by discussing social influence chapter in P.Young's recent paper "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence and Social Learning". The analysis assumes a population which is infinitely large and encounters between individuals are purely random. The issue that Young looks at in this paper that few have analysed in the literature is the *speed* of adoption of innovations in a threshold model.

Young considers a population where each agent has a social threshold r_i , such that when a proportion r_i of the population have adopted, agent i will also want to adopt the innovation. Young also has a level of inertia in his model such that even though an agent's social threshold has been passed, they will now only adopt the innovation at a rate $\lambda > 0$.

Importantly Young does not want to restrict the distribution of thresholds and as such considers a general cumulative distribution function of thresholds, $F(r)$. In order to create initial diffusion there must be some agents who will adopt the innovation even though no-one else has, and so $F(r) > 0$. There must be some of these genuine innovators in order to "get the ball rolling", once they adopt then others (with low r_i) may be persuaded to adopt and so on in a domino type effect. The first interesting question is whether this dynamic will continue until everyone adopts, or whether the process will stop with only some people adopting.

In order to address this question Young considers at some time point t , a proportion $F(p(t)) > 0$ have had their thresholds passed, of which $p(t)$ have adopted, and so

$$\dot{p}(t) = \lambda[F(p(t)) - p(t)], \lambda > 0$$

Of which the inverse function

$$\forall x \in [0, b), t = p^{-1} = (1/\lambda) \int_0^x \frac{dr}{F(r) - r}$$

can be obtained. By considering the first fixed point of F such that $F(b) = b$, it can be seen that as $x \rightarrow b$, the right hand side of the later equation goes to infinity due to $\frac{1}{F(r)-r}$, and as such $p(t)$ must converge to b in the long run.

This occurs when the number of agents whose thresholds have been crossed becomes close to those who have already adopted, thus essentially bringing the process to a halt apart from those taking up the innovation due to inertia alone.

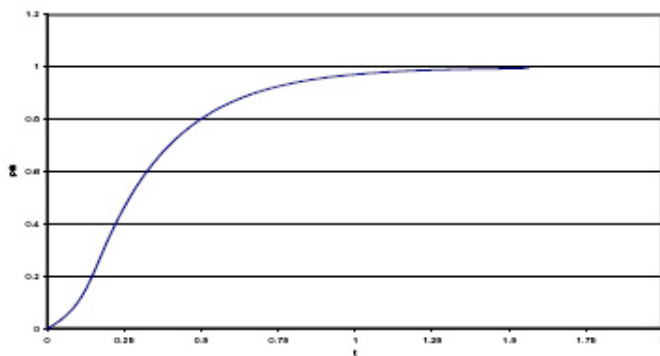
And so an innovation will only be adopted by everyone in the population if the first fixed point of the process occurs at $F(1) = 1$. As in Granovetter and other papers if a population has many people willing to adopt only once there has been a small uptake then the innovation may not get off the ground due to an early fixed point.

The second and more important issue that Young investigates is the shape of the adoption curve given $F(r)$, with particular interest in the speed of adoption. The essence of this investigation comes from the equation for the *acceleration* of adoption, which Young forms by differentiating the first equation detailed and dividing by $\dot{p}(t)$ to give

$$\frac{\ddot{p}(t)}{\dot{p}(t)} = \lambda[f(p(t)) - 1]$$

And so one can see that if $f(p(0)) > 1$ then the process will initially be accelerating. Consider also that $F(p(x))$ is increasing for x in the neighborhood of the origin. Then if both these conditions are met, we can see that the process will experience a period of acceleration. Indeed Young shows the adoption curve will exhibit super-exponential growth for some

time, which is faster than an exponential growth rate as $\frac{\ddot{p}(t)}{\dot{p}(t)}$ is increasing. Young also shows that this result holds when the inertia rate λ is heterogeneously distributed. A diagram taken from Young's paper of such growth is shown by the blue line in the adoption curve below:



This extreme growth occurs because of two reasons. Firstly, a significant proportion of initial agents adopting means that other agents are willing to adopt the innovation. Secondly, when $F(p(x))$ is increasing for x in the neighborhood of the origin there are simultaneously more agents willing to adopt than there were initially, creating a double effect.

Interestingly it is the mass of those initially willing to innovate who drive and determine the speed and shape of adoption. If there are enough agents willing to adopt when no-one else has, and enough agents when only a few have adopted, then the innovation will experience a period of extreme growth. If the initial innovators are not present, despite many others potentially willing to take up the innovation later on, the innovation may never get off the ground.

However some criticism may come due to Young's analysis concentrating on the initial growth of adoption and the final adoption levels separately. From the analysis given it seems possible that the adoption curve may exhibit super-exponential growth initially but the innovation may not be

taken up by a large proportion of the population. Similarly, an adoption curve which initially decelerates may in the long term reach 100% take-up. Although one can deduce the final presence of the innovation in individual cases by determining the first fixed point of $F(p(x))$, some discussion on the link between initial growth of adoption and the final adoption levels between would be welcomed.

In this chapter Young has provided some insight into the speed to adoption of new innovations and not just the long term equilibrium solutions. Interestingly he shows that the shape of the adoption curve is highly dependent on the nature of the innovators of a population.

5.8 Social network thresholds in the diffusion of innovations - Valente 1996

In this paper Valente conducts an empirical investigation on the diffusion of innovations in three separate cases. The three cases are the take up of doctors prescribing a new medical drug in four US communities in the mid 1950s, the diffusion of hybrid corn among 692 Brazilian farmers in 11 villages in the 1940s and 1950s, and finally the adoption of modern family planning in by 1000 Korean women in 1973. Valente is able to obtain real estimates for individual thresholds at which innovation adoption occurs and then analyse the diffusion of the innovation in the population as a whole.

A key point is that Valente categorises possible reasons for an individual's adoption into two separate sources. The first is influence from social networks. Each individual was asked to name certain people who influenced their decision making with regards to the relevant issue, thus determining a social network each agent was part of. Some individuals with many nominations can then be considered as leaders. An individual's *adoption threshold* is determined by the number of adopters in their network at the time they themselves adopt the innovation. Each individual is categorised according to their network threshold as either very low, low, high or very high.

The second source of diffusion is that of external influence from society on an individual independent of their social network. Such influences may include the proportion of adopters in society as a whole as well as media and advertising. Each individual in the study has a quantifiable score of external influence determined by the number of medical journals subscribed to, the number large Brazilian towns visited annually or the

exposure to family planning advertising, for the three separate cases. Each individual is categorised by the time at which they adopt the innovation, independent of their social network, as either early adopters, early majority, late majority or laggards, and I shall describe this category as an individual's *adoption time*. These four categories will be compared to those of adoption thresholds, such that an individual who has a very low threshold and is also an early adopter will be described as being consistent across groups.

The results from the paper then come from comparing individual's adoption threshold with their adoption time, whilst also looking at external exposure and leadership distributions. It is possible for an individual to have a low adoption threshold yet be in the late majority of adopters if no-one in their social network adopts the innovation early on. Similarly an individual can have a high adoption threshold yet be an early adopter if many people in their social network adopt the innovation early on.

The first results state that 43% of doctors are in the same groups for adoption threshold and adoption time, 47% of Brazilian farmers and 64% of Korean women. We also see that the laggards are almost entirely composed of very low or very high threshold individuals, those who are just not exposed to the innovation in their network or just plain stubborn.

The second set of results examines the individual external influence scores with respect to adoption thresholds and adoption time. Firstly we see that for the majority of cases early adopters have the highest external influence score, suggesting that external influence leads to their highly innovative behaviour. Furthermore external influence appears to affect adoption time far more than adoption thresholds. For a given threshold level, an individual's adoption time is always earlier the greater their ex-

ternal influence score, but this is not the case with thresholds. Instead external influence scores are highest when an individual's threshold level is consistent with their adoption time grouping. We also see that laggards with low threshold levels generally have a low external influence score further implying isolation as the reason for non-adoption. Valente also points out that specific external influence results across the three groups are not statistically similar, in particular medical journal subscription had little correlation with an individual's adoption threshold or adoption time.

The final set of results compare the individual leadership scores with respect to adoption thresholds and adoption time. Interestingly those with the highest leadership scores are again generally those individuals whose threshold level is consistent with their adoption time grouping. This implies that individuals who are consistent in their adoption decision are popular role models. We also see particularly large inconsistencies across the three groups. In Korea the women with high leadership scores tended to be those who were early adopters suggesting those early adopters may have high relative education and therefore social status. However, in the medical example the doctors in the late majority with high thresholds tended to have the highest leadership scores, implying a link between conservativeness and leadership. No clear pattern in leadership emerged from the Brazilian case.

In summary the majority of individuals are consistent in their adoption threshold and their adoption time, and these individuals are often the leaders in their communities. We can also see that external influence leads the way for the first individuals to adopt in society in general yet does not have a strong effect on an individual having a low social network threshold, instead those with consistent adoption threshold and adoption

time groupings have the most external influence. And so it would appear that consistency is the key issue, with those individuals having both medium adoption thresholds and adoption times, and also having the high external influence levels often being the leaders in their respective societies.

5.9 Conclusion

This literature review studies an array of models based upon dynamics created by individuals who possess a social threshold, such that they can be persuaded to take an action if enough of their own population already take the action. The literature has produced many interesting models and results on issues such as social integration, rioting, contagion and the diffusion of innovations.

Within the literature there exists a variety of results from models with different assumptions. In our opinion a critical point of the whole literature is whether it is assumed that the threshold distribution can change over time and how this takes place. Without this assumption populations can often get stuck in low adoption equilibria, yet without it any society can reach high uptake levels. Whether populations engage in high levels of adoption is obviously a key point and so I believe more attention should be devoted to the nature of changing threshold distributions.

Amongst the literature we also see that few models have populations which are subject to any level of perturbations, and therefore the robustness of some results could be put into question. Indeed Granovetter shows that his results significantly change when he introduces some noise into his model. As such I feel future research should be based both upon the nature of evolving thresholds distributions and models which are subject

to small perturbations.

References

- Dodds, Peter Sheridan, and Duncan J. Watts. 2004. "Universal Behavior in a Generalized Model of Contagion." *Physical Review Letters*, 92 (21): 218701-1 to 218701-4.
- Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology*, 83: 1420-1443.
- Granovetter, Mark, and Roland Soong. 1988. "Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence." *Sociological Methodology*, 18: 69-104.
- Macy, Michael W. 1991. "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review*, 56: 730-747.
- Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1: 143-186.
- Valente, Thomas W. 1996. "Social Network Thresholds in the Diffusion of Innovations." *Social Networks* 18: 69-89.
- P. Young 2008. "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning," *American Economic Review*, forthcoming.