Competing Conventions*

Philip R. Neary†

November 9, 2010

JOB MARKET PAPER

Abstract

This paper studies a new coordination game, the Language Game, of a large but finite population. The population is partitioned into two groups of identical agents. Each player shares a common two-action strategy set and interacts pairwise with everyone else. Both symmetric profiles are pareto-efficient strict equilibria, but the groups rank them differently. The profile where successful coordination occurs only within-group, with each group adopting their most preferred action, is also an equilibrium provided the smaller group's preferences are sufficiently strong. In all dynamically stable long run outcomes, players in the same group adopt the same action. Three properties, that do not matter for equilibrium selection in the homogeneous agent models of Kandori, Mailath, and Rob (1993) and Young (1993), do matter in the Language Game. These are: group size, preference over alternative equilibria, and rates of group adaptiveness ("group dynamism"). A relative increase in group dynamism is always weakly beneficial.

*I would like to thank Sarada, Nageeb Ali, Prashant Bharadwaj, Tiffany Chou, Vince Crawford, Silke Januszewski Forbes, Dalia Ghanem, Roger Gordon, Youjin Hahn, Oana Hirakawa, Michihiro Kandori, Chulyoung Kim, Daniel Lima, Stephen Morris, Paul Niehaus, Frances Ruane, Bill Sandholm, Lucas Siga, James Slevin, Hee-Seung Yang, and Peyton Young for helpful comments. I would especially like to thank Joel Sobel for all his help. Remaining errors are mine.

†*Email*: pneary@ucsd.edu; *Web*: http://econ.ucsd.edu/∼pneary/; *Address*: Department of Economics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

"*Nobody will ever win the battle of the sexes.*
There's just too much fraternizing with the enemy."

- Henry Kissinger

1 Introduction

A large population coordination problem is one wherein all parties can realize mutual gains, but only by making mutually consistent decisions. Often, such mutually consistent decisions require that everybody behave identically. For example: writing papers in English is a must if that is the conventional medium; it makes sense to buy a MAC if all your friends already own one; driving on the opposite side of the road as the oncoming traffic hardly seems wise; etc.

The emergence of coordinated outcomes in large societies, referred to as *conventions* by Lewis (1969) and Young (1993, 1996, 2001), has been studied using the framework of evolutionary game theory. In the canonical model, players are drawn from a homogeneous population and randomly matched to play a symmetric 2×2 game of pure coordination. This situation is then repeated with players assumed to follow some simple updating rule, that allows population behaviour to be tracked in a manageable way. However, by construction, the canonical model can only be used for studying the emergence of conventions in societies where all agents have the same preferences. This is limiting, since in many situations of interest, like the adoption of technological standards (Katz and Shapiro, 1985; Farrell and Saloner, 1985; Arthur, 1989), people often have different tastes, and so what might be best for some may not be best for all.

In this paper, I propose a new strategic situation, the "Language Game", that allows the study of conventions in a heterogeneous environment. The population is composed of not one, but two homogeneous groups. Each player has the same two-action strategy set and interacts pairwise with everybody else. I assume that successful coordination is good, while all types of failed coordination are bad. Precisely, a player's utility is linearly increasing in the number of others who adopt the same action. Each player has a most preferred coordinated outcome - the key feature is that these most preferred outcomes differ across the groups.

Uniform adoption of either action is always a strict equilibrium (convention) of the Language Game. If the smaller group has sufficiently strong preferences for one action over the other, then the profile where members of each group adopt their preferred action and forfeit coordination with those of the other group, can also be supported as an equilibrium. Such an outcome highlights one positive difference between my model and existing ones.

Next, the Language Game becomes the stage game of a repeated interaction. Time is discrete, begins at $t = 0$, and goes forever. Payoffs are received every period and actions for tomorrow must be chosen at the end of today. Following Kandori, Mailath, and Rob (1993) (hereafter KMR), players are assumed (*i*) to be *myopic*, and thus behave as though the following period is the last, and (*ii*) to exhibit some *inertia*, so that not everybody reacts instantaneously to his/her environment. Whatever individual best responses do get made in a given period are aggregated to obtain deterministic (nonergodic) group dynamics. Any decentralized adjustment process with these features will lock in on some convention with probability one.

The concept of *stochastic stability* (Foster and Young, 1990) enables crisp predictions to be made about long run behaviour. The basic idea is that players occasionally, and independently, choose non-optimal responses. These continual "mistakes" or "mutations" perturb the dynamics in such a way that population behaviour now spends the bulk of time in the neighbourhood of only certain *selected* conventions. Such equilibrium selection in the Language Game depends on three factors: group size, group payoffs, and how fast each group adapts.[1] Increased rate of adapting and increased numbers are always more likely to bring about a group's most preferred outcome. However, stronger preferences need not always be desirable.

In the canonical model with only one homogeneous population, for any convention each player's behaviour and payoff are identical. As such, the literature has focused primarily on the tension that arises when the "good" pareto-dominant equilibrium action does not coincide with the "safe" risk-dominant one. Foster and Young (1990), KMR and Young (1993) were the first to show that evolutionary forces coupled with mutations will propel population behaviour towards risk-dominance.[2] While the defini-

[1] Rates of adapting, or "group dynamism", permits many interpretations. It can be thought of as how sensitive on average a particular group is to their surroundings, or as the frequency of death and replacement, or even as adjustment costs varying across groups since some find change less burdensome.

[2] This negative result has generated a vast literature. The following papers show that the pareto-dominant equilibrium action can emerge in similar settings. Ely (2002), Oechssler (1999) and Oechssler (1997) are models with endogenous pairwise interactions, Canals and Vega-Redondo (1998) and Robson and Vega-Redondo (1996) vary the frequency with which players may interact, while Kim and Sobel

tion of risk-dominance, and hence of a risk dominant equilibrium, is not so clear in the Language Game, the stochastically stable equilibria are never pareto-inefficient. However, due to the existence of multiple pareto-efficient strict equilibria, not all welfare measures rank conventions identically. So inefficient outcomes may emerge although the nature of the inefficiency is different.

Ensuing work showed that the classic risk-dominance selection result is robust to situations where players interact with only small sets of neighbours, rather than with the population at large (see Ellison (1993), Blume (1993), Young (2001), and Peski (2010)). It is quite a startling finding that altering network architecture cannot in any way influence equilibrium selection.[3] However, this is simply a consequence of the fact that all pairwise interactions are the same, which implies that each individual's coordination problem is sufficiently similar, and hence in the presence of neighbours who occasionally make mistakes so is optimal behaviour, i.e. choose the risk dominant action. The Language game is a situation with more than one type of pairwise interaction, and an immediate implication of moving it to a network, is that equilibrium selection is highly sensitive to network topology. This issue is studied in a companion paper Neary (2010a).

The plan of the paper is as follows. In the next Section, I present a simple example that demonstrates how homogeneous groups with heterogeneous preferences can be a more natural way to think about certain large population situations, and can provide previously unexplored insights. The Language Game is formally defined in Section 3, where I also characterize the set of pure strategy equilibria. Section 4 shows how decisions at the individual level are aggregated to yield group dynamics, and illustrates via some examples how path dependence may be influenced by the dynamics. This analysis is carried forward to Section 5 which contains the main results on equilibrium selection. Section 6 looks at welfare properties of the selected equilibria, while Section 7 examines how the set of selected equilibria varies as Language Game parameters change. Section 8 concludes and discusses some potential avenues for future work.

(1995) add a round of costless communication, "cheap talk", before actions are taken.

[3] Ellison (1993) did note that network architecture can dramatically affect that speed at which selection will occur, and further notes that if selection takes a long time to occur then perhaps the validity of evolutionary forces should be called in to question.

2 A Story

The story is an extension of one from KMR. I begin by reminding the reader of theirs and then building on it. There is a university dormitory of 10 identical students, referred to as Group A. Each Group A student uses a computer system s chosen from the set $\{a, b\}$. Each evening, the students assemble in the study hall, where everybody encounters everyone else. When two students interact, they can collaborate by playing games, sharing files, swapping add-ons, etc. But - and this is key - meetings are fruitful if and only if both students use the same computer system. Assume that system a is inherently superior to system b. This induces the following local-interaction pure coordination game, G^{AA},

		A_{II}: a	A_{II}: b
A_I	a	$2, 2$	$0, 0$
	b	$0, 0$	$1, 1$

where A_I and A_{II} are any pair of Group A students. G^{AA} has two pure strategy equilibria, (a, a) and (b, b), and a third equilibrium in mixed strategies, where each player puts weight $1/3$ on a. The population coordination problem has two pure strategy equilibria in which all 10 students adopt a common system, a or b. These profiles are denoted by $\mathbf{a}$ and $\mathbf{b}$ respectively.

It is assumed, again following KMR, that students occasionally have the opportunity to change their computer, and that students are myopic in that decisions are taken based on the current distribution of computers. This generates *Darwinian* dynamics, in which population behaviour is always *drifting* towards either $\mathbf{a}$ or $\mathbf{b}$. Initial conditions are key: if more than one third of the population begins using system a (4 or more since the population is of size 10), then outcome $\mathbf{a}$ will be reached, while if 7 or more students start out using b, then $\mathbf{b}$ will be the final resting point. The reasoning is simple, all players collectively agree on what action is a best response, so the best response today must be at least as good a response tomorrow as the number of players taking that action can only (weakly) increase.

However, when *trembles* or *mistakes* or *experimentation* are incorporated into the dynamics, it is possible to *select* between strict equilibrium outcomes. Suppose that the probability that a student mistakenly chooses the computer that is not an optimal response is given by ε. It takes 4 or more simultaneously occurring mistakes to dislodge the system from **b**, and 7 or more to get away from **a**. The most likely events of this form occur with probability of orders ε^4 and ε^7 respectively. For small values of ε, $\varepsilon^7 \ll \varepsilon^4$, and so KMR conclude that when agents are myopic best responders, who occasionally make mistakes, that outcome **a** is far more likely to be observed in the long run.

A key component of the above story was that system a is *inherently superior* to system b. While in many coordination problems it is plausible to believe that coordination on one particular strategy is better (by any metric) than another, words like "better" derive from primitive preferences, and preferences are individual by nature. In a population with heterogeneous agents, what might be best for some may not be best for all.

To illustrate the impact of adding heterogeneity, consider the following extension to the above story. Suppose Group A are "slackers" - they must also do coursework on their machines, but their main use for computers is playing games. Instead of assuming that computer system a is flat out better than b, let us suppose that system a more readily supports gaming platforms, which justifies Group A's underlying preference for coordinating on it. Suppose further that there is another dorm of 5 more students in the next building. This dorm is unconnected to the first dorm, and I shall call those in this dorm, Group B. Every night these 5 Group B students meet in a separate study hall and exchange software, etc. Again, interactions are beneficial if and only if the pair involved share the same system. However, those in Group B are more "serious" scholars, and platform b suits their scholarly needs better. The local-interaction between two Group B students is given by the following pure coordination game, G^{BB},

		B_{II}: a	B_{II}: b
B_I	a	$1, 1$	$0, 0$
	b	$0, 0$	$2, 2$

where B_I and B_{II} are any two Group B students. By an identical analysis to that given above for the Group A coordination problem, left to their own devices Group B

will adopt computer system b in the long run.[4] Now, consider the 15 person population as a whole. Writing a group-symmetric profile[5] as a vector, $(\mathbf{s}_A, \mathbf{s}_B)$, with the Group A profile written first, there are 4 strict equilibria: $(\mathbf{a}, \mathbf{a}), (\mathbf{a}, \mathbf{b}), (\mathbf{b}, \mathbf{a}), (\mathbf{b}, \mathbf{b})$. One can easily see, and it is quite intuitive, that the long run population profile will be $(\mathbf{a}, \mathbf{b})$. So with both groups isolated, each group internally coordinates on its preferred outcome.

The situation changes when the dorms are connected. Suppose that in an effort to free up space, the university stipulates that both groups should study in the larger Group A study room which can accommodate 5 extra bodies. (This frees up the smaller study room for other activities.) So now, all 15 students meet in the same room every evening. Everybody interacts with everybody else, and within-group local-interactions are as before. It remains to specify the local-interaction that occurs when students from opposite groups meet. This is described by the coordination game G^{AB}, in which the row player, A_i, is from Group A, while the column player, B_j, is a Group B student,

		B_j: a	B_j: b
A_i	a	$2,1$	$0,0$
A_i	b	$0,0$	$1,2$

G^{AB} is not symmetric. It has two pareto-efficient pure strategy equilibria, (a, a) and (b, b), over which players' preferences disagree. Considering the new coordination problem, interactions are now occurring both within- and across-group. Each Group A student interacts with 9 fellow Group A students and 5 Group B students, while each Group B student interacts with 4 other Group B students and 10 Group A students. The only group-symmetric equilibria to this new situation are $(\mathbf{a}, \mathbf{a})$ and $(\mathbf{b}, \mathbf{b})$.[6] While both are pareto-efficient, the 10 Group A students prefer $(\mathbf{a}, \mathbf{a})$, while the 5 Group B students view $(\mathbf{b}, \mathbf{b})$ as most desirable.

I now pause and ask the reader to predict what they think the long run outcome will be (recalling that $(\mathbf{a}, \mathbf{a})$ and $(\mathbf{b}, \mathbf{b})$ are the only viable candidates). One conjecture might be the following. Even though behaviour evolves in a decentralised manner

[4]The probability of Group B transitioning from $\mathbf{a}$ to $\mathbf{b}$ occurs with probability of order ε^2, while that of transitioning from $\mathbf{b}$ to $\mathbf{a}$ occurs with probability of order ε^4.

[5]A *group-symmetric* profile is one in which those in the same group take the same action. A symmetric profile is group-symmetric, though plainly the reverse need not be true.

[6]While $(\mathbf{a}, \mathbf{b})$, and $(\mathbf{b}, \mathbf{a})$ were equilibria when the two groups were unconnected, they are no longer equilibria when all students study together.

via individual best responses, group preferences are *mirrored* so the greater Group A numbers should somehow collectively force its preferred outcome, $(\mathbf{a}, \mathbf{a})$, onto the population at large.

It turns out that the answer is subtle and depends on a variety of factors aside from group sizes and group preferences. The first of these is mistakes. In the original story with only Group A, all agents were identical so assuming they all make time- and profile-independent mistakes with equal probability seemed not completely unreasonable. However, with a population composed of two types of agents, if Group A students tremble with probability ε^A while Group B students tremble with probability ε^B, there is no obvious reason to conclude that $\varepsilon^A = \varepsilon^B$.

The second complicating factor is a property I refer to as *group dynamism*. Group dynamism can be thought of as the rate at which a group responds. It may be that Group A students are more lethargic than Group B students. Perhaps on average only one Group A student updates his action in a given period, whereas all Group B students update their action every period. In the real world, there is no reason to suppose that different groups respond at identical rates. In fact, often they do not.

However, with these caveats in mind, let's begin by assuming that: (1) payoffs are as given in G^{AA}, G^{AB} and G^{BB}; (2) the group sizes are 10 and 5 for Groups A and B respectively; (3) mistakes are such that $\varepsilon^A = \varepsilon^B = \varepsilon$; and (4) each period, both groups evolve according to the best-reply dynamic in which all students react. Now, let us calculate how easy it is to dislodge the population from each symmetric profile.

Suppose first that the population profile is $(\mathbf{b}, \mathbf{b})$. Any Group B student needs to see a minimum of 10 others taking action a in order to switch his/her action, while a Group A student needs to see a minimum of 6. Let's say between 6 and 9 students accidentally chose action a (it is not important how these students are distributed across the two groups). In the following period, all Group B students choose action b, but all Group A students choose action a, so that the new profile is $(\mathbf{a}, \mathbf{b})$. With no further mistakes, all students take action a the following period. The conclusion is that 6 or more simultaneous mistakes are sufficient to shift the population from $(\mathbf{b}, \mathbf{b})$ to $(\mathbf{a}, \mathbf{a})$.

Now assume that the current profile is $(\mathbf{a}, \mathbf{a})$. Payoffs are mirrored so the 6-player and 10-student bounds are still relevant. If between 6 and 9 mistakes occur whereby students accidentally choose action b, it is enough to induce the 5 Group B students to take action b, but not enough to induce the 10 Group A students to do so. Thus, next period's profile is $(\mathbf{a}, \mathbf{b})$, and with no further trembles, the system reverts to $(\mathbf{a}, \mathbf{a})$. It can be computed that a minimum of 10 simultaneous mistakes are required to shift the system from $(\mathbf{b}, \mathbf{b})$ to $(\mathbf{a}, \mathbf{a})$.

Transitioning from $(\mathbf{a}, \mathbf{a})$ to $(\mathbf{b}, \mathbf{b})$ requires an event that occurs with probability of order ε^{10}, while transitioning from $(\mathbf{b}, \mathbf{b})$ to $(\mathbf{a}, \mathbf{a})$ one with probability of order ε^{6}. For small values of ε, the second transition is far more likely. So provided that: (i) everybody interacts with everybody else; (ii) both groups respond according to best-reply dynamics; (iii) the probability of a student making a mistake is small, equal for all students, and independent of the current population profile; and (iv) payoffs and group-sizes are as specified above. Then, the informal analysis concludes that the unique long run outcome will be $(\mathbf{a}, \mathbf{a})$.

Now suppose that the groups adapt at different rates. Precisely: (1) payoffs are as given in G^{AA}, G^{AB} and G^{BB}; (2) the group sizes are 10 and 5 for Groups A and B respectively; (3) mistakes are such that $\varepsilon^A = \varepsilon^B = \varepsilon$; and (4) each period, Group B evolves according to the best-reply dynamic, while only one Group A student best responds. Again let us calculate how easy it is to dislodge the population from each of the equilibria. The bounds of 6 and 10 are derived from preferences and group sizes, not dynamics, so those have not changed. The difference in this analysis is that it will matter exactly who is making mistakes.

This time, start with the population at $(\mathbf{b}, \mathbf{b})$. Suppose 6 Group A students mistakenly choose a (it does matter that these 6 students are from Group A). Next period, Group B students maintain taking action b, while one more Group A student adopts action a, so that the total number taking action a is increased to 7. The following period 8 Group A students are using action a, and so on. Once all 10 Group A students are taking action a, action a becomes optimal for Group B students who all immediately adopt it. Thus, 6 of the "right kind" of mistakes are enough to transition from $(\mathbf{b}, \mathbf{b})$ to $(\mathbf{a}, \mathbf{a})$.

Now let the current profile be $(\mathbf{a}, \mathbf{a})$, and suppose that 6 Group A students accidentally choose action b, (again, it matters that these 6 students are from Group A). At this new profile, Group preferences disagree. The reactiveness of the groups means that the following period, all 5 Group B students adopt action b, while one of the 6 Group A students who mistakenly chose action b reverts back to action a. Thus, there are 10 $\big(= 5 + (6 - 1)\big)$ students taking action b. This is enough for all Group B students to maintain action b and for Group A students to prefer action b. From this new situation, by an analysis similar to the previous paragraph, the behaviour of the population moves incrementally to $(\mathbf{b}, \mathbf{b})$.

Under these different dynamics, transitioning from $(\mathbf{b}, \mathbf{b})$ to $(\mathbf{a}, \mathbf{a})$ still occurs with probability of order ε^6. More importantly however, the likelihood of transitioning from $(\mathbf{a}, \mathbf{a})$ to $(\mathbf{b}, \mathbf{b})$ has been lowered to ε^6. So provided that: (*i*) everybody interacts with everybody else; (*ii*) Group B best-replies, while Group A is more lethargic; (*iii*) the probability of a student making a mistake is small, equal for all students, and independent of the current state; and (*iv*) payoffs and group-sizes are as specified above. Then, the informal analysis concludes that both symmetric outcomes are equally likely to be observed in the long run.

So what, the reader might ask, is the point of this section? There is certainly no hint of a crisp result like the risk-dominance prediction of KMR and Young (1993).[7] But in fact, the lack of a clean result is precisely the point. That is, what prediction results I do obtain, are incredibly fragile. That group size and strength of payoffs affect long run behaviour is intuitive but not predicted by homogeneous agent case. In that framework, once an equilibrium is risk-dominant it is selected (even though it may be pareto-dominated to an arbitrary extent), and this is of course independent of population size.

Once we move to the details of the dynamics the situation becomes even worse. It is well known (Bergin and Lipman, 1996) that any strict equilibrium may be selected with appropriately defined mutations. However, the risk-dominance result is robust to both uniform errors and also payoff dependent errors (Blume, 1993). This is not the case for the Language Game described above - though given that strength of payoffs affect selection even under uniform errors, it is unsurprising that payoff dependent dynamics yield different selection results.

[7] I am applying existing selection techniques to a new strategic situation, and in this situation it is unclear how concepts like risk-dominance ought be defined. One might be tempted to suggest that the profile $(\mathbf{a}, \mathbf{b})$ is risk-dominant, but for the parameters given above this profile is not an equilibrium. Thus we would be in the unusual situation of having the risk-dominant profile be unstable.

3 The Language Game

3.1 The Model

The Language Game, $\mathcal{G}$, is defined as the tuple $\{\mathcal{N}, \Pi, S, \mathbb{G}\}$, where $\mathcal{N} := \{1, \ldots, N\}$ is the *population* of players; $\Pi := \{A, B\}$ is a *partition* of $\mathcal{N}$ into two nonempty homogeneous *groups* A, B of sizes N^A, N^B (≥ 2) respectively; $S := \{a, b\}$ is the set of actions common to all players; $\mathbb{G} := \left\{G^{AA}, G^{AB}, G^{BB}\right\}$ is the collection of pairwise local-interactions, where G^{AA} is the exchange that occurs whenever a player from Group A meets a player from Group A, etc. G^{AA}, G^{AB}, and G^{BB} are given as follows,

$$G^{AA} \qquad\qquad A_2 \qquad\qquad A_1\ \begin{array}{c|c|c|} & a & b \\ \hline a & p, p & 0, 0 \\ \hline b & 0, 0 & 1-p, 1-p \\ \hline \end{array}$$

$$G^{BB} \qquad\qquad B_2 \qquad\qquad B_1\ \begin{array}{c|c|c|} & a & b \\ \hline a & 1-q, 1-q & 0, 0 \\ \hline b & 0, 0 & q, q \\ \hline \end{array}$$

$$G^{AB} \qquad\qquad B_j \qquad\qquad A_i\ \begin{array}{c|c|c|} & a & b \\ \hline a & p, 1-q & 0, 0 \\ \hline b & 0, 0 & q, 1-p \\ \hline \end{array}$$

The Language Game is a simultaneous move game, in which players do not randomize. Utilities are the sum of payoffs earned from playing the field, where the same action must be used with one and all.[8] I assume that $p, q \in (1/2, 1)$, so Group A members prefer to coordinate on a, and Group B prefer to coordinate on b. Even though within-group local-interactions are constrained symmetric while those across-group are not, note that all local-interactions are *opponent independent* in that a player's payoff depends only on the actions chosen and not the other player's identity.[9] Thus a player

[8]While this has a different interpretation to a game of random matching, strategic behaviour is the same in both.

[9]This is obviously a gross simplification and may not be applicable for many real world situations. See Neary (2010b) for extensions of this model to situations where not only the value of successful coordination may be opponent dependent, but even what might be the optimal action with one Group may not be optimal with another, e.g. coordination 'vs' anti-coordination.

cares only about the number of others expected to choose the same action, and not on who those others are.

With only two types of agents, population behaviour can be written concisely. Let $[\omega]_K$ denote the number of players in group $K \in \Pi$ using action a. Call $\omega = \big([\omega]_A , [\omega]_B \big)$ the *state* of the play, where the *state space* is $\Omega := \big\{0, \ldots, N^A\big\} \times \big\{0, \ldots, N^B\big\}$. For any state $\omega \in \Omega$, define $n_a = [\omega]_A + [\omega]_B$, and $n_b = N - n_a$. The utility a player in group $K \in \Pi$ receives from taking action $s \in \{a, b\}$ in state ω, written $U^K(s;\omega)$, is given by

$$U^A(a;\omega) := \Big(n_a - 1\Big)p \tag{1}$$

$$U^A(b;\omega) := \Big(N - n_a - 1\Big)(1-p) \tag{2}$$

$$U^B(a;\omega) := \Big(n_a - 1\Big)(1-q) \tag{3}$$

$$U^B(b;\omega) := \Big(N - n_a - 1\Big)q \tag{4}$$

Before discussing individual behaviour, I should mention genericity. Letting $\mathbb{N} := \{1, 2, \ldots\}$, the set of Language Games can be parameterized by $\Theta = \{(N^A, N^B, p, q) : N^A, N^B \in \mathbb{N}\backslash \{1\} ; p, q \in (1/2, 1)\}$. For a given game $\mathcal{G} \in \Theta$ and a given group $K \in \Pi$, the statement "if there does not exist a state $\omega \in \Omega$ such that $U^K(a;\omega) = U^K(b;\omega)$", will be abbreviated to "*genK*". If there exists such a state, the shorthand is "*ngenK*". The subset of the parameter space for which any indifference occurs can easily be shown to have a closure of measure zero, and so when both genA and genB, following standard terminology I say $\mathcal{G}$ is *generic*. Otherwise, $\mathcal{G}$ is *nongeneric*.

3.2 Individual Behaviour

Let $\mathbb{R}$ denote the real line, and $\mathbb{R}_+$ its positive part. For any $x \in \mathbb{R}$, let $\lceil x \rceil := \min \{n \in \mathbb{N} \,|\, x \leq n\}$ and $\lfloor x \rfloor := \max \{n \in \mathbb{N} \,|\, x \geq n\}$. Define the following,[10]

$$n_a^A := \min \Big\{n_a \,\Big|\, U^A(a;\omega) > U^A(b;\omega)\Big\} = \lceil (1-p)N + (2p-1) \rceil, \tag{5}$$

$$n_a^B := \min \Big\{n_a \,\Big|\, U^B(a;\omega) > U^B(b;\omega)\Big\} = \lceil q(N-2) + 1 \rceil, \tag{6}$$

$$n_b^A := \min \Big\{n_b \,\Big|\, U^A(a;\omega) < U^A(b;\omega)\Big\} = \lceil p(N-2) + 1 \rceil, \tag{7}$$

$$n_b^B := \min \Big\{n_b \,\Big|\, U^B(a;\omega) < U^B(b;\omega)\Big\} = \lceil (1-q)N + (2q-1) \rceil \tag{8}$$

[10]The values given are for generic $\mathcal{G}$. For the relevant nongeneric case, increase the values by 1.

n_a^A is the number of players taking action a for a player from Group A to strictly prefer action a, etc. When $p = q$, by symmetry $n_a^A = n_b^B$ and $n_b^A = n_a^B$. This appears to suggest that the strategic situation is mirrored for groups A and B. While this is true at the individual level, it need not be true at the level of the group.

Let $\Omega^{A,a\succ b}$ and $\Omega^{B,a\succ b}$ denote the set of states such that A players and B players respectively *strictly* prefer action a to action b. Similarly define $\Omega^{A,b\succ a}$, and $\Omega^{B,b\succ a}$.

$$\Omega^{A,a\succ b} := \left\{\omega \in \Omega \,\middle|\, [\omega]_A + [\omega]_B \geq n_a^A\right\} \tag{9}$$

$$\Omega^{B,a\succ b} := \left\{\omega \in \Omega \,\middle|\, [\omega]_A + [\omega]_B \geq n_a^B\right\} \tag{10}$$

$$\Omega^{A,b\succ a} := \left\{\omega \in \Omega \,\middle|\, [\omega]_A + [\omega]_B \leq N - n_b^A\right\} \tag{11}$$

$$\Omega^{B,b\succ a} := \left\{\omega \in \Omega \,\middle|\, [\omega]_A + [\omega]_B \leq N - n_b^B\right\} \tag{12}$$

Sets $\Omega^{A,a\succeq b}$, $\Omega^{B,a\succeq b}$, $\Omega^{A,b\succeq a}$, and $\Omega^{B,b\succeq a}$ are defined likewise but for *weak* preference. Generically, these sets of weak- and strict-preference coincide. Letting $\subseteq$ ($\subset$) denote weak (strict) inclusion, we have the following lemma whose simple proof is omitted.

Lemma 1. *When $N^A \geq 2$ and $N^B \geq 2$,*

1. $\Omega^{B,a\succeq b} \subseteq \Omega^{A,a\succ b}$
2. $\Omega^{A,b\succeq a} \subseteq \Omega^{B,b\succ a}$

An immediate implication of Lemma 1 is that, even non-generically, there is no state such that members of both groups are simultaneously indifferent, i.e. ngenA and ngenB cannot both hold.

The interpretation of the lemma is as follows. Fix a state ω. If all members of Group B (A) weakly prefer action a (b) at this particular state, then this same action is the unique best response for all members of Group A (B), and hence is a best response for the population as a whole. It does <u>not</u> say that if action a (b) is preferred by Group A (B), it must simultaneously be preferred by Group B (A). That is, there may exist a state such that group preferences differ. The following provides mild sufficient conditions for the existence of such a state.

Lemma 2. *If either*

- *N is even, or*

- *N is odd and $N > 2 + \frac{1}{p+q-1}$.*

then,

$$\Omega^{A,a \succ b} \cap \Omega^{B,b \succ a} \neq \emptyset$$

Proof. The proof is found in Appendix C. □

3.3 Equilibria

Behaviour at states $(0,0),(0,N^B),(N^A,0)$, and (N^A,N^B), is referred to as *group-symmetric* for obvious reasons. These four states will appear repeatedly throughout the paper, and are denoted by ω_{bb},ω_{ba}, ω_{ab}, and ω_{aa} respectively. For a given game, $\mathcal{G}$, let $E(\mathcal{G})$ denote the set of group-symmetric equilibria. With an abuse of terminology, I will refer to $E(\mathcal{G})$ as the *equilibrium set*, since it turns out (Section 4) that group-symmetric equilibria are the only serious candidates for long run behaviour. The following Theorem, stated without proof, classifies $E(\mathcal{G})$ for various parameters.

Theorem 1. *In the Language Game, $\mathcal{G}$,*

1. *State ω_{aa} is always a strict equilibrium.*

2. *State ω_{bb} is always a strict equilibrium.*

3. *State ω_{ba} is never an equilibrium.*

4. *State ω_{ab} is an equilibrium if and only if*

$$p \geq \frac{N^B}{N-1} \quad \text{and} \quad q \geq \frac{N^A}{N-1}$$

Parts 1 - 2 of Theorem 1 are easily understood, since deviating from a symmetric profile means failing to coordinate with everyone in the population. Part 3 is also very simple. One of the groups must be (weakly) smaller, and at state ω_{ba}, members of this (weakly) smaller group observe strictly more than half the players in the population adopting their preferred action. Hence they wish to deviate.

The intuition for part 4 is as follows. State ω_{ab} involves each player successfully coordinating on their most preferred action with only those from his/her own group. To sustain ω_{ab} as an equilibrium, the high payoffs earned from within-group interactions, must exceed those that could be earned from successful coordination with the members of the other group on a less preferred action. This requires the product of "own group size" and "preferred local-interaction payoff" be sufficiently large for each player. That is, a player must either be part of the larger group, or part of a group with a strong relative preference for one action over the other, or both. The inequality for the larger group clearly always holds, and so one must only check that of the smaller group.

While Theorem 1 is simple, it is also intuitive. The following two examples, which are carried throughout the paper, illustrate precisely why. They further demonstrate that while both symmetric profiles are socially efficient, members of the different groups prefer different ones. This observation stimulates the discussion of welfare in Section 6.

Example 1. Let $\mathcal{G}^1 = (10, 5, 3/5, 2/3)$.
For these parameters ω_{ab} is not an equilibrium, because from this state Group B players have an incentive to deviate to action a. Precisely, the second inequality of part 4 in Theorem 1 does not hold. The reader can check that the first inequality of part 4 in Theorem 1, relevant to Group A players, does hold, as it must since Group A is the larger group.

Example 2. Let $\mathcal{G}^2 = (10, 5, 3/5, 5/6)$.
Group B members now have a stronger relative preference for coordinating on action b over action a. As in Example 1, the first inequality of Theorem 1 part 4, for ω_{ab} to be an equilibrium holds (there is no need to recheck as parameters relevant to Group A are the same as they were in $\mathcal{G}^1$). This time however, the inequality relevant to Group B members also holds. While group B still has only half as many members as group A, the relative reward for coordinating on action b over action a for group B members has increased sufficiently that even coordinating with a small number of players on their most preferred action can compensate for the larger number of *failed* coordinations.

4 Evolutionary Dynamics

4.1 Specification

Now suppose the Language Game becomes the stage game of a repeated interaction. Time is discrete, begins at $t = 0$, and goes forever. Utilities are received every period and actions for tomorrow are chosen at the end of today. I avail of precisely the assumptions placed on individual behaviour from KMR's evolutionary model.

Assumption 1. *Inertia: At the end of each period, a nonempty subset of players are provided with the opportunity to revise their strategy for the following period.*

Assumption 2. *Myopia: When a player does react, he best responds to the current environment.*

One possible explanation put forward for high inertia is that in many situations changing an action is a costly excercise. Another is that players observe only slices of information so their understanding of the game may be imperfect, and that this may cause them to stand by the status quo for longer than might be optimal. Myopia usefully captures the notion that players are boundedly rational and/or do not understand the dynamics of the population at large. Furthermore, it follows quite naturally for systems with high inertia, since in this case, a best response against the current population profile should not only generate a high payoff tomorrow, but also a "pretty good" payoff for some time in near the future.

Aggregating responses from the individual level to the population level, the evolution of population behaviour may be described by a deterministic dynamic, Ψ, where

$$\begin{aligned}\omega_{t+1} &:= \Psi(\omega_t)\\ &= \big(\Psi^A(\omega_t), \Psi^B(\omega_t)\big)\end{aligned}$$

The mappings $\Psi^A : \Omega \to \{0, \ldots, N^A\}$ and $\Psi^B : \Omega \to \{0, \ldots, N^B\}$ are the respective Group dynamics. Following from myopia, strategic decisions are made by looking back at today's environment, and making a choice of action for tomorrow, based on what would have been an ideal strategy to have held earlier today. However, since there is inertia, perhaps not all agents myopically best respond every period. Thus, each of Ψ^A and Ψ^B possess the "*Darwinian*" property of KMR, so that Ψ satisfies the following definition.

Definition 1 (Group-Darwinian Adjustment Process)**.** Say that $\Psi = (\Psi^A, \Psi^B)$ has the *Group-Darwinian Property* if, for any $K \in \{A, B\}$,

1. for all $\omega \notin \{\omega' \mid [\omega']_K = 0, N^K\}$,

$$\operatorname{sign}\big(\Psi^K(\omega) - [\omega]_K\big) = \operatorname{sign}\big(U^K(a;\omega) - U^K(b;\omega)\big) \tag{13}$$

2.
   - $\Psi^K(\omega) = 0$, if $[\omega]_K = 0$ and $U^K(a;\omega) \leq U^K(b;\omega)$
   - $\Psi^K(\omega) = N^K$, if $[\omega]_K = N^K$ and $U^K(a;\omega) \geq U^K(b;\omega)$

Group-Darwinianism naturally extends the Darwinian property of KMR to a situation with multiple groups. It is *similar*, in that it makes the standard evolutionary assumption that better strategies are no worse represented next period. But it is *different* since, (i) for certain states, groups may disagree over which strategies are "better" (Lemma 2), and (ii) the rate at which each group's best responses are better represented next period is left unspecified, and need not be the same across groups.

As time proceeds the population reacts every period, and so interest lies in repeated applications of Ψ. For all ω, let $\Psi^0(\omega) = \omega$, and for $m \in \mathbb{N}$ define the m-fold repetition of Ψ, Ψ^m, inductively as $\Psi^m(\omega) = \Psi\big(\Psi^{m-1}(\omega)\big)$. Define the set of *rest points*, $\Omega_0 := \{\omega \mid \Psi(\omega) = \omega\} \subset \Omega$. It can easily be shown that for each $\omega \in \Omega$, there exists $\hat{\hat{m}}(\omega)$, such that for all $m \geq \hat{\hat{m}}(\omega)$, $\Psi^m(\omega) \in \Omega_0$. Finally, let $\hat{m} := \max_\omega \hat{\hat{m}}(\omega)$, and for each $\omega \in \Omega_0$, define the *basin of attraction* of ω by $\mathcal{V}(\omega) := \{\omega' \mid \forall\, m \geq \hat{m}, \Psi^m(\omega') = \omega\}$. Generically, $\Omega_0 = E(\mathcal{G})$, and so the state space can be partitioned into $\{\mathcal{V}(\omega)\}_{\omega \in E(\mathcal{G})}$. When ngenA, it is possible that $(n_a^A - 1, 0) \in \Omega_0$, and when ngenB, it may be that $(N^A, N^B - n_b^B + 1) \in \Omega_0$.

When Lemma 2 holds, the basins of attraction depend on the exact specification of Ψ. Keeping track of these is of prime concern for issues of equilibrium selection in Section 5. To assist in this, define a partial ordering on Ω as follows. If ω and ω' are elements of Ω, write $\omega \geqslant_a \omega'$, if $[\omega]_A \geq [\omega']_A$ and $[\omega]_B \geq [\omega']_B$. That is, $\omega \geqslant_a \omega'$ if, in state ω, there are (weakly) more players in both groups taking action a. The pair $(\Omega, \geqslant_a)$ is a complete lattice with *bottom* element ω_{bb} and *top* element ω_{aa}.[11]

While the Group-Darwinian property of Definition 1 seems appealing at first, the class of dynamics satisfying it is still too broad to show general results. The following additional constraint placed on the dynamics will make tracking of population behaviour easier.

[11] Lattices are briefly discussed in Appendix A.

Definition 2 (Monotonic Adjustment Process)**.** Say that $\Psi : \Omega \to \Omega$ is *monotonic*, if for any pair $\omega', \omega'' \in \Omega$,

$$\omega' \geqslant_a \omega'' \Rightarrow \Psi(\omega') \geqslant_a \Psi(\omega'') \tag{14}$$

It is easy to construct adaptive processes satisfying Definition 1, but not Definition 2, and vice versa. I limit attention to the class of dynamics satisfying both Definitions, and refer to these to as *monotone Group-Darwinian* processes. Soon I place further restrictions on this class, but first, a useful result.

Lemma 3. *For any monotonic Group-Darwinian process, and any state $\omega \in \Omega_0$, the set $\mathcal{V}(\omega)$ is convex.*

Proof. The proof is contained in Appendix C. □

The tractable subclass of monotonic Group-Darwinian dynamics on which I focus, are those where each group responds at a constant rate. The relative rates at which groups adapt is a property I term *group-dynamism*.

Definition 3. Say that Group $K \in \{A, B\}$ responds at *constant-rate-ξ_K*, if there exists $\xi_K \in \left\{1, \ldots, N^K\right\}$, such that for all $\omega \in \Omega$, and all t,

$$\Psi^K(\omega_t) = [\omega_{t+1}]_K = \begin{cases} \max\left\{0, [\omega_t]_K - \xi_K\right\}, & \text{if } U^K(a;\omega_t) < U^K(b;\omega_t), \\ [\omega_t]_K, & \text{if } U^K(a;\omega_t) = U^K(b;\omega_t), \\ \min\left\{[\omega_t]_K + \xi_K,\ N^K\right\}, & \text{if } U^K(a;\omega_t) > U^K(b;\omega_t), \end{cases}$$

Definition 4. Say that Ψ is a *constant rate dynamic* if both groups adapt at constant rates. If Groups A and B respond at *constant-rates* ξ_A and ξ_B respectively, write $\Psi = (\Psi^A_{\xi_A}, \Psi^B_{\xi_B})$.

Constant rate dynamics have a simple interpretation. Next period, a fixed number of new agents from each group adopt that group's best response (provided this new number, when added to the original number of agents who were already taking that action, does not exceed the size of the group). The *best-reply* dynamic, $\mathcal{B} := (\mathcal{B}^A, \mathcal{B}^B)$, is a constant rate dynamic with $\xi_A = N^A$ and $\xi_B = N^B$. I now define what it means for one group to be *more dynamic* than the other. Clearly, best-replying will be the most reactive a group can be.

Definition 5. If Ψ is a constant rate dynamic, say that Group $K \in \{A, B\}$ is *(weakly) more-dynamic* than Group $K' \neq K$, written $\Psi^K \succ_d (\succeq_d)\Psi^{K'}$, if

$$\xi_{K'} < (\leq) \min\Big\{\xi_K\,,\, N^{K'}\Big\}$$

Groups adapt at the same rate, $\Psi^A \sim_d \Psi^B$, if neither is more dynamic. Formally, $\Psi^A \sim_d \Psi^B$ if either (i) $\xi_A = \xi_B = \xi \in \big\{1, \ldots, \min\big\{N^A - 1, N^B - 1\big\}\big\}$, or (ii) one of the following holds: $\big(\xi_A = N^A \leq \xi_B \leq N^B\big)$ or $\big(\xi_B = N^B \leq \xi_A \leq N^A\big)$. The first condition says that neither group best-replies, and, whenever possible, next period equal numbers of new agents from each group adopt that group's best response. The second says that if the smaller of the two groups is best-replying, then even if, whenever possible, the larger group has more agents reacting each period, both groups are still said to be evolving at the same rate.

Armed with Definitions 3-5, it will now be possible to make positive statements about varying dynamics. Normative statements are more problematic. While increased adaptiveness is a desirable property in the Language Game (Section 5), that is because locally risk-dominant actions coincide with most-preferred equilibrium actions. If these actions did not accord, then greater group dynamism could be detrimental.[12]

4.2 Path Dependence

In KMR, dynamics are defined on a linear state space, and when the common local-interaction is a game of pure coordination, generically there are two rest points, one at either end. KMR emphasize that *path dependence* of population behaviour rests crucially on the initial conditions, but that the final outcome is "*independent ... to all but the coarsest features of the dynamics*". The intuition for this was discussed in the story of Section 2. Loosely, once the process starts heading in a particular direction, it cannot "turn around". However, in the Language Game, the final outcome can depend not only on the initial state, ω_0, but also on the exact specification of the dynamics.

To illustrate how path dependence may be sensitive to both the initial conditions and also the specifics of the dynamics, recall Example 1 from Section 3, where $\mathcal{G}^1 = \big\{(N^A, N^B), (p, q)\big\} = \{(10, 5), (3/5, 2/3)\}$. Figure 1 shows the state space Ω as an

[12] That is, to be part of a less dynamic ("lazier") group may be desirable in certain situations. Thanks to Frances Ruane for originally pointing this out.

11×6 lattice, with $[\omega]_A \in \{0, \ldots, 10\}$ on the horizontal-axis, and $[\omega]_B \in \{0, \ldots, 5\}$ on the vertical-axis. Each state is depicted by a circle. The set of blue circles is $\Omega^{A,b \succeq a}$, while those red circles comprise $\Omega^{B,a \succeq b}$. These sets are defined by $(n_a^A, n_b^A, n_a^B, n_b^B) = (7, 9, 10, 6)$, calculated using equations 5-8. At the states depicted by hollow circles, $\Omega^{A,a \succ b} \cap \Omega^{B,b \succ a}$, group preferences disagree (by Lemma 1, Group A prefers a, Group B prefers b).

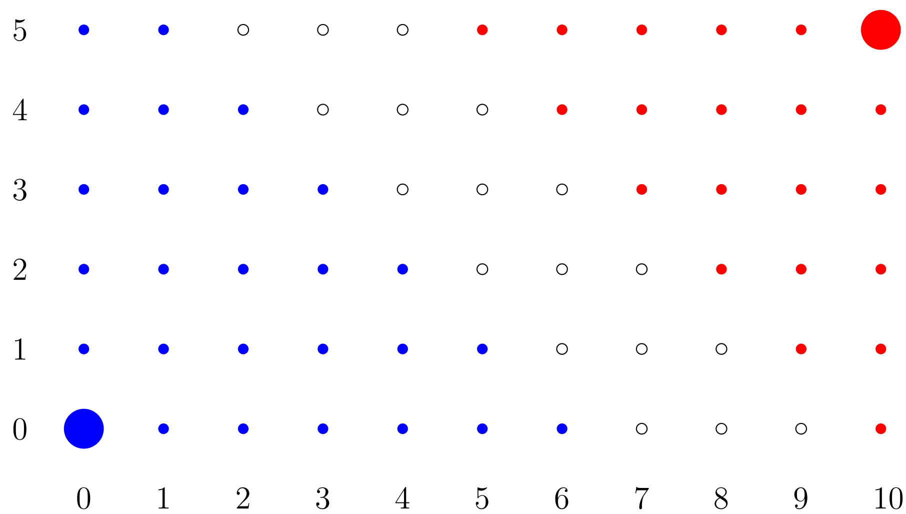

Figure 1: $\Omega^{A,b \succeq a}$ - $\Omega^{A,a \succ b} \cap \Omega^{B,b \succ a}$ - $\Omega^{B,a \succeq b}$.

A large circle denotes a rest point. Corner states ω_{bb} and ω_{aa} are always rest points, while corner state ω_{ab} is a rest point if the conditions of Theorem 1 part 4 are satisfied. While states $(n_a^A - 1, 0)$ and $(N^A, N^B - n_b^B + 1)$ can be rest points non-generically, no other state can be. Example 1 is generic and the conditions of Theorem 1 part 4 are not satisfied, and so in Figure 1, the only two rest points are $\omega_{bb} = (0, 0)$ and $\omega_{aa} = (10, 5)$.

To understand Group-Darwinianism, consider the state $(5, 3)$. Since $(5, 3) \in \Omega^{A,a \succ b} \cap \Omega^{B,b \succ a}$, it must be that $\Psi\big((5, 3)\big) \in \{6, \ldots, 10\} \times \{0, 1, 2\}$. While monotonicity is not shown in Figure 1, it is also easily understood when coupled with Group-Darwinianism. Consider the pair of states, $(5, 2)$ and $(5, 3)$. Clearly $(5, 3) \geqslant_a (5, 2)$, and so $\Psi\big((5, 3)\big) \geqslant_a \Psi\big((5, 2)\big)$. Since $(5, 2) \in \Omega^{A,a \succ b} \cap \Omega^{B,b \succ a}$, $\Psi\big((5, 2)\big) \in \{(6, \ldots 10\} \times \{0, 1\}$. Now suppose that $\Psi\big((5, 2)\big) = (8, 1)$. Then $\Psi\big((5, 3)\big)$ is further restricted to lie in the set $\Big\{(8, 1), (9, 1), (10, 1), (8, 2), (9, 2), (10, 2)\Big\}$.

Figure 1 is a preference map, with $\big\{\Omega^{A,b \succeq a}, \Omega^{B,a \succeq b}, \Omega^{A,a \succ b} \cap \Omega^{B,b \succ a}\big\}$ as the partition. To partition Ω into $\{\mathcal{V}(\omega)\}_{\omega \in \Omega_0}$, further information on the details of the dynamics are needed. To see how these details can matter, it is instructive to start by looking at behaviour at individual states for varying dynamics.[13]

Example 3 $(\Psi = (\mathcal{B}^A, \mathcal{B}^B))$**.** From Figure 1, the interpretation is easy: no matter what the current state, the following state must be ω_{bb}, ω_{ab}, or ω_{aa}. Blue states jump immediately to ω_{bb}, black states to ω_{ab}, and red states to ω_{aa}. Formally, for any $\omega \in \Omega^{A,b\succeq a}$, $\Psi(\omega) = \mathcal{B}(\omega) = \omega_{bb}$; for any $\omega \in \Omega^{B,a\succeq b}$, $\Psi(\omega) = \mathcal{B}(\omega) = \omega_{aa}$; and for any $\omega \in \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a}$, $\mathcal{B}(\omega) = \omega_{ab}$ and $\mathcal{B}^2(\omega) = \omega_{aa}$. That is, states in $\Omega^{A,a\succ b} \cap \Omega^{B,b\succ a}$ transition first to ω_{ab}, and from there to ω_{aa}, and so with $\Psi = \mathcal{B}$, are considered red.

Example 4 $(\Psi = (\Psi_1^A, \mathcal{B}^B))$**.** Figure 2 illustrates the basins of attraction for this scenario. As before, states in $\Omega^{B,a\succeq b}$ are denoted by solid red circles, and states in $\Omega^{A,b\succeq a}$ by solid blue circles. States for which the groups have conflicting preferences, $\Omega^{B,b\succ a} \cap \Omega^{A,a\succ b}$, are again denoted by hollow circles. However, it is not the case that $\Omega^{B,b\succ a} \cap \Omega^{A,a\succ b} \subseteq \mathcal{V}(\omega_{aa})$, as it was when $\Psi = \mathcal{B}$. Hollow red circles eventually lead to $\Omega^{B,a\succeq b}$ and hence to ω_{aa}, but hollow blue circles lead immediately to $\Omega^{A,b\succeq a}$ and eventually to ω_{bb}.

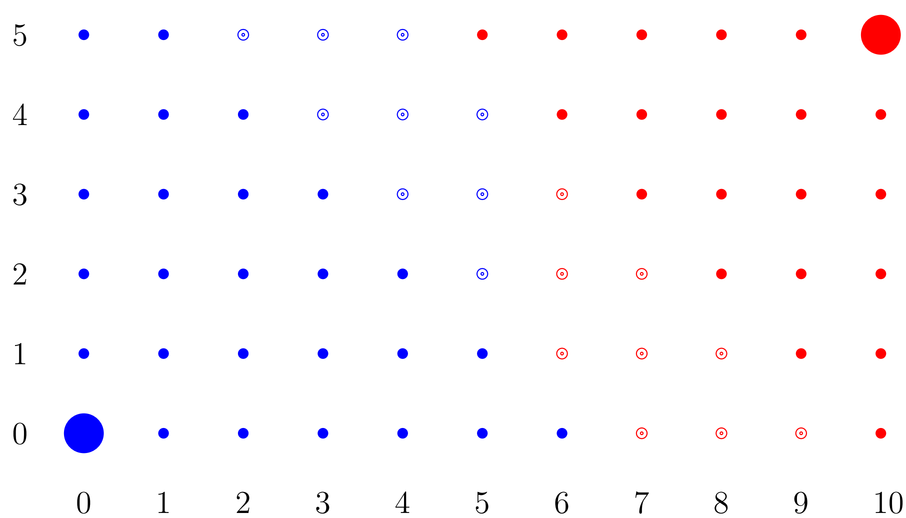

Figure 2: $\mathcal{V}(\omega_{bb})$ and $\mathcal{V}(\omega_{aa})$ when $\Psi = (\Psi_1^A, \mathcal{B}^B)$.

To be clear what is happening, again consider state $(5,3)$. When $\Psi = (\mathcal{B}^A, \mathcal{B}^B)$ as in Example 3, the dynamics terminate at $(10,5)$ via the path $\{(5,3) \rightarrow (10,0)$

[13] The formal analysis of how basins of attraction vary across constant rate adaptive processes is carried out in Appendix B.

$\rightarrow (10,5)\}$. When $\Psi = (\Psi_1^A, \mathcal{B}^B)$ as in Example 4, $(5,3)$ leads to $(0,0)$, via the path $\{(5,3) \rightarrow (6,0) \rightarrow (5,0) \rightarrow (4,0) \rightarrow (3,0) \rightarrow (2,0) \rightarrow (1,0) \rightarrow (0,0)\}$.

We can also trace how the boundaries of the basins of attraction vary.[14] When $\Psi^A \sim_d \Psi^B$, all basins of attraction, and hence their boundaries, are the same.[15] By Lemma 3, for any monotone dynamics the basins will be convex. This is easily seen by inspection of Figures 1 and 2. From Figure 1, and for any Ψ such that $\xi_A = \xi_B$,

- $\big(\mathcal{V}(\omega_{bb})\big)_+ = \big(\mathcal{V}(\omega_{bb})\big)_{++} = \{(1,5),(2,4),(3,3),(4,2),(5,1),(6,0)\}$
- $\big(\mathcal{V}(\omega_{aa})\big)_- = \big(\mathcal{V}(\omega_{aa})\big)_{--} = \{(2,5),(3,4),(4,3),(5,2),(6,1),(7,0)\}$

while in Figure 2, with $\Psi = (\Psi_1^A, \mathcal{B}^B)$,

- $\big(\mathcal{V}(\omega_{bb})\big)_+ = \{(4,5),(5,4),(6,0)\}$
- $\big(\mathcal{V}(\omega_{bb})\big)_{++} = \{(4,5),(5,4),(5,3),(5,2),(5,1),(6,0)\}$
- $\big(\mathcal{V}(\omega_{aa})\big)_- = \{(5,5),(6,1),(7,0)\}$
- $\big(\mathcal{V}(\omega_{aa})\big)_{--} = \{(5,5),(6,4),(6,3),(6,2),(6,1),(7,0)\}$

5 Equilibrium Selection

Any deterministic Group-Darwinian dynamic, Ψ, induces a time-homogeneous Markov process on the finite state space Ω. Let P be the associated Markov matrix, where for every pair of states $\omega', \omega'' \in \Omega$, $P(\omega', \omega'') \geq 0$ denotes the probability of transitioning from ω' to ω'', and for each $\omega \in \Omega$, $\sum_{\omega'} P(\omega, \omega') = 1$.

For any finite set X, let $\triangle(X)$ denote the set of distributions on X. A stationary distribution of P is a row-vector $\mu \in \triangle(\Omega)$, such that $\mu P = \mu$. The set of stationary distributions is denoted $\triangle_0(\Omega)$. Writing $\text{supp}(\mu)$ for the support of μ, say that $D \subset \Omega$ is a recurrent class, if for all $\omega \in \Omega$, and all $\mu \in \triangle_0(\Omega)$ with $\text{supp}(\mu) \subset D$, $\mu(\omega) > 0 \iff \omega \in D$. A state is *recurrent* if it is contained in a recurrent class, and *transient* otherwise. A singleton recurrent class is an *absorbing state*.

[14] The boundaries, both upper (lower) and total upper (total lower), are defined in Appendix A.

[15] In fact, this statement holds true under the weaker condition that $\xi_A \geq n_a^A$ and $\xi_B \geq n_b^B$.

All Markov processes possess at least one stationary distribution, while *ergodic* Markov processes possess only one. The third assumption of KMR, perturbs the deterministic dynamics in such a way as to induce a new Markov process that is ergodic.[16]

Assumption 3. *Behavioural Mutation: There is a small probability that an agent may choose an action at random.*

One interpretation is as follows. After afforded decisions have been taken, but before payoffs are made, with probability $\varepsilon^A > 0$ ($\varepsilon^B > 0$) each Group A (B) player switches his current action choice, and with probability $1 - \varepsilon^A$ ($1 - \varepsilon^B$) maintains his action.[17]

Even for constant rate dynamics, there is no grounds for always assuming $\Psi^A \sim_d \Psi^B$. Similarly, there is no reason to suppose that behavioural mutations occur with equal likelihood for members of different groups.[18] So while interest will lie in the case where $(\varepsilon^A, \varepsilon^B) \to (0,0)$, I will be making the strong assumption that $\varepsilon^A = \varepsilon^B = \varepsilon$, for all states and all time periods. It is tempting to insist on a milder condition like $\varepsilon^A = O(\varepsilon^B)$ and $\varepsilon^B = O(\varepsilon^A)$,[19] but the selection results may differ.[20]

For a given $\varepsilon > 0$, the above perturbations define a new ergodic Markov process with associated transition matrix P^ε, and unique stationary distribution μ^ε. By continuity, the accumulation point of $\{\mu^\varepsilon\}_{\varepsilon>0}$, $\mu^\star$, is a stationary distribution of $P := \lim_{\varepsilon\downarrow 0} P^\varepsilon$. Our interest lies in the states to which $\mu^\star$ assigns positive probability.

Definition 6. State ω is *stochastically stable* if $\mu^\star(\omega) > 0$, and *uniquely stochastically stable* if $\mu^\star(\omega) = 1$. Let $\Omega^\star$ denote the set of stochastically stable states.

Write $\mathcal{L}$ for the collection of recurrent classes. For the Language Game, it can be shown that $\mathcal{L} = \{\{\omega\}\}_{\omega\in\Omega_0}$, where Ω_0 is the set of rest points as defined in Section 4.

[16] Technically this assumption makes the process *irreducible* and *aperiodic* which for finite state Markov processes is sufficient for ergodicity. See Karlin and Taylor (1975).

[17] This is a different interpretation to that given in KMR, but it generates the same switching probabilities.

[18] Nor is there any reason to suppose mutations are both state- and time-independent. The effects that subtle differences in mutation rates can have on equilibrium selection are examined in Bergin and Lipman (1996).

[19] Letting both ε^A and ε^B be functions of ε, we say that $\varepsilon^A(\varepsilon) = O\big(\varepsilon^B(\varepsilon)\big)$ as $\varepsilon \downarrow 0$, if and only if there exists positive numbers M and δ, such that $\varepsilon^A(\varepsilon) \leq |M\varepsilon^B(\varepsilon)|$ for all $\varepsilon < \delta$.

[20] However, this should not be muddled with the results of Bergin and Lipman (1996) (see footnote 18 above). To my knowledge, there are no games for which the Ising model dynamics of Blume (1993) select *different* long run equilibria to the dynamics of KMR. In Neary (2010a), I show how these two dynamics can arrive at different selection results for an open set of parameters for the Language Game.

We can express $\Omega^\star$ as the union of recurrent classes as follows

$$\exists\, \mathcal{M} \subset \mathcal{L} \text{ such that } \Omega^\star = \bigcup_{D \in \mathcal{M}} D$$

Calculating $\mu^\star$ is the objective. This is done using tree-surgery techniques from Freidlin and Wentzell (1998), first introduced to game theory in Foster and Young (1990). To do so, it will be useful to view states in Ω as the vertices of a fully connected directed graph, $\Gamma^\star$. An edge in $\Gamma^\star$ from ω' to ω'' is denoted $(\omega' \to \omega'')$. A *walk* from ω' to ω'' is a sequence of edges $\{(\omega_i \to \omega_{i+1})\}_{i=0}^{m-1}$ where $\omega_0 = \omega'$, and $\omega_m = \omega''$. A path is a walk in which the vertices are distinct. A typical path from ω' to ω'' is denoted by $h(\omega', \omega'')$, and the set of all paths from ω' to ω'' by $H(\omega', \omega'')$. Extending this, the set of all paths from a state ω to a set $Q \not\ni \omega$ can be defined as follows, $H(\omega, Q) = \cup_{\omega' \in Q} H(\omega, \omega')$.

Following KMR, we assume that for any pair ω', ω'', the following limit exists

$$\begin{aligned} c_\Psi(\omega', \omega'') &:= \lim_{\varepsilon \downarrow 0} \frac{\log P^\varepsilon(\omega', \omega'')}{\log \varepsilon} \\ &= \|\Psi(\omega'), \omega''\| \end{aligned}$$

where $c_\Psi : \Omega \times \Omega \to \mathbb{R}_+ \cup \{\infty\}$ is a *cost function*. The value it takes for a particular pair (ω', ω'') is interpreted as the minimum number of simultaneous mutations needed to transition directly from ω' to ω'', or in graph theoretic terms, as the cost of edge $(\omega' \to \omega'')$ in $\Gamma^\star$.[21]

For any function $\tau : \Omega \to \Omega$, a path from ω' to ω'' in τ, is a path $\{(\omega_0 \to \omega_1), (\omega_1 \to \omega_2), \ldots, (\omega_{m-1} \to \omega_m)\}$, where $\omega_0 = \omega'$ and $\omega_m = \omega''$, such that $\tau(\omega_i) = \omega_{i+1}$ for all $i = 0, \ldots, m-1$. An ω-tree, τ_ω, is a mapping $\tau_\omega : \Omega \to \Omega$ such that: (i) $\tau_\omega(\omega) = \omega$; (ii) for every $\omega' \in \Omega \backslash \{\omega\}$, there is a unique path in τ_ω from ω' to ω. Say that ω'' is a *successor* of ω' in τ_ω if $\tau_\omega^m(\omega') = \omega''$ for some $m \geq 1$, and the *immediate successor* if $m = 1$.

For each ω, T_ω is the set of all ω-trees. The cost of ω-tree, $\tau_\omega \in T_\omega$, is the sum of

[21] Note that c_Ψ is allowed to take the value ∞. This could be the case if a transition is impossible under the dynamics. In this paper, the range of c_Ψ is the range of $\|,\|$ which is $\{0, \ldots, N\}$.

the costs of its edges,

$$c_\Psi(\tau_\omega) = \sum_{\omega' \neq \omega} c_\Psi\big(\omega', \tau_\omega(\omega')\big)$$

For the Language Game, $\mathcal{G}$, and cost function, c_Ψ, define the set of states that achieve minimum cost ω-trees as

$$\Xi(\mathcal{G}, c_\Psi) := \left\{\omega^\star \in \Omega \,\middle|\, \text{for any } \omega \in \Omega, \min_{\tau_{\omega^\star} \in T_{\omega^\star}} c_\Psi(\tau_{\omega^\star}) \leq \min_{\tau_\omega \in T_\omega} c_\Psi(\tau_\omega)\right\}$$

The following is the result of Freidlin and Wentzell (1998). Note it's relation to Definition 6 above.

Lemma 4. *State ω is stochastically stable if $\omega \in \Xi(\mathcal{G}, c_\Psi)$, and uniquely stochastically stable if $\{\omega\} = \Xi(\mathcal{G}, c_\Psi)$. That is, $\Xi(\mathcal{G}, c_\Psi) = \Omega^\star$.*

By Young (1993), Theorem 4, the stochastically stable states are contained in a recurrent class. We can therefore restrict attention to minimum cost ω-trees of recurrent states. For the Language Game, $\mathcal{L} = \{\{\omega\}\}_{\omega \in \Omega_0}$, and $\mathcal{V}(\omega') \cap \mathcal{V}(\omega'') = \emptyset$ for all distinct $\omega', \omega'' \in \Omega_0$. Thus, the key to computing ω-trees of the absorbing states, is to find a path of minimum cost from each absorbing state to the convex basin of attraction of the others. For any pair of states $\omega', \omega'' \in \Omega$, denote by $c^\star_\Psi(\omega', \omega'')$ the cost of the path of minimum cost between them. That is,

$$c^\star_\Psi(\omega', \omega'') := \min_{\{(\omega_j \to \omega_{j+1})\}_{j=0}^{n-1} \in H(\omega', \omega'')} \sum_{m=0}^{n-1} c_\Psi(\omega_m, \omega_{m+1})$$

The two main results in this Section concern equilibrium selection. Before presenting these however, the following Lemma is needed. It says that a path of minimum cost out of a region of the state space in which the dynamics are unambiguous, involves a direct transition out. An immediate and important consequence is that it holds for the symmetric profiles ω_{bb} and ω_{aa}.

Lemma 5. *Let Ψ be a monotonic Group-Darwinian adjustment process. Then,*

1. *For all $\omega \in \Omega^{A,b \succeq a}$, the minimum of $c_\Psi(h')$ over all paths $h' \in H\big(\omega, \Omega \backslash \Omega^{A,b \succeq a}\big)$ is attained by*

$$h^\star := \Big\{(\omega \to \omega^\star)\Big\}$$

where

$$\omega^\star \in \operatorname*{argmin}_{\hat{\omega} \in (\Omega \backslash \Omega^{A,b \succeq a})_-} \|\Psi(\omega), \hat{\omega}\|$$

2. *For all $\omega \in \Omega^{B,a\succeq b}$, the minimum of $c_\Psi(h'')$ over all paths $h'' \in H\big(\omega, \Omega\backslash\Omega^{B,a\succeq b}\big)$ is attained by*

$$h^{\star\star} := \Big\{(\omega \to \omega^{\star\star})\Big\}$$

where

$$\omega^{\star\star} \in \operatorname*{argmin}_{\hat{\omega} \in (\Omega \backslash \Omega^{B,a \succeq b})_+} \|\Psi(\omega), \hat{\omega}\|$$

However, the paths $h^\star$ and $h^{\star\star}$ above attaining these minimum costs need not be unique.

Proof. The proof is found in Appendix C. □

Generically $\Omega^{A,b\succeq a} \subseteq \mathcal{V}(\omega_{bb})$ and $\Omega^{B,a\succeq b} \subseteq \mathcal{V}(\omega_{aa})$. So all that remains is to classify behaviour for states in $\big(\Omega^{A,a\succ b} \cap \Omega^{B,a\succ b}\big)$. It turns out that the behaviour of the dynamics in these states can be key for selection. When $E(\mathcal{G}) = \{\omega_{bb}, \omega_{aa}\}$, the set $\big(\Omega^{A,a\succ b} \cap \Omega^{B,a\succ b}\big)$ has more of an effect than when $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$. The analysis of each case is quite different and so are looked at separately. The second case is easier to begin with.

5.1 Equilibrium set is $\{\omega_{\mathbf{bb}}, \omega_{\mathbf{ab}}, \omega_{\mathbf{aa}}\}$

We begin by calculating the minimum cost ω-trees of each convention, when the groups adapt at constant and equal rates.

Theorem 2. *Suppose Condition 4 of Theorem 1 holds, so $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, and that the monotonic Group-Darwinian adjustment process is such that both groups evolve at constant and equal rates. Let $\tau^\star_{\omega_{bb}}$, $\tau^\star_{\omega_{ab}}$, and $\tau^\star_{\omega_{aa}}$, denote minimum cost ω-trees for ω_{bb}, ω_{ab}, and ω_{aa} respectively. Then,*

$$c_\Psi(\tau^\star_{\omega_{bb}}) = n^B_b + n^A_b - N^B \tag{15}$$

$$c_\Psi(\tau^\star_{\omega_{ab}}) = n^B_b + n^A_a \tag{16}$$

$$c_\Psi(\tau^\star_{\omega_{aa}}) = n^A_a + n^B_a - N^A \tag{17}$$

The set of stochastically stable states are those with ω-tree of minimum cost. That is,

$$\Xi(\mathcal{G}, c_\Psi) = \underset{\omega \in \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}}{\operatorname{argmin}} c_\Psi(\tau^\star_\omega)$$

The details are found in Appendix C. Here I discuss the intuition. The proof rests on computing paths of minimum cost between the six pairs of coventions, $(\omega_{bb}, \omega_{ab})$, $(\omega_{bb}, \omega_{aa})$, $(\omega_{ab}, \omega_{bb})$, $(\omega_{ab}, \omega_{aa})$, $(\omega_{aa}, \omega_{bb})$, and $(\omega_{aa}, \omega_{ab})$. By Lemma 5, the minima of $c_\Psi(h')$ over all $h' \in H(\omega_{bb}, \omega_{ab})$, and of $c_\Psi(h'')$ over all $h'' \in H(\omega_{aa}, \omega_{ab})$, are attained by

$$\begin{aligned}
h^\star(\omega_{bb}, \omega_{ab}) &= \big\{\big(\omega_{bb} \to (n_a^A, 0)\big)\big\} \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \,|\, \omega' = \Psi^m\big((n_a^A, 0)\big) \text{ for some } m \geq 0\Big\} \\
h^\star(\omega_{aa}, \omega_{ab}) &= \big\{\big(\omega_{aa} \to (N^A, N^B - n_b^B)\big)\big\} \\
&\qquad \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \,|\, \omega' = \Psi^m\big((N^A, N^B - n_b^B)\big) \text{ for some } m \geq 0\Big\}
\end{aligned}$$

Next, use Lemma 8 to show that the minima of $c_\Psi(h')$ over all $h' \in H(\omega_{ab}, \omega_{bb})$, and of $c_\Psi(h'')$ over all $h'' \in H(\omega_{ab}, \omega_{aa})$, are attained by

$$\begin{aligned}
h^\star(\omega_{ab}, \omega_{bb}) &= \big\{\big(\omega_{ab} \to (n_b^A - N^B, 0)\big)\big\} \\
&\qquad \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \,|\, \omega' = \Psi^m\big((n_b^A - N^B, 0)\big) \text{ for some } m \geq 0\Big\} \\
h^\star(\omega_{ab}, \omega_{aa}) &= \big\{\big(\omega_{ab} \to (N^A, n_a^B - N^A)\big)\big\} \\
&\qquad \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \,|\, \omega' = \Psi^m\big((N^A, n_a^B - N^A)\big) \text{ for some } m \geq 0\Big\}
\end{aligned}$$

Lastly, note that when both groups adapt at constant and equal rates, $\mathcal{V}(\omega_{ab})$ is "sandwiched" between $\mathcal{V}(\omega_{bb})$ and $\mathcal{V}(\omega_{aa})$. That is, for all $\omega' \in \mathcal{V}(\omega_{bb})$ and $\omega'' \in \mathcal{V}(\omega_{aa})$ with $\omega' \leqslant_a \omega''$, there exists $\hat{\omega} \in \mathcal{V}(\omega_{ab})$ such that $\omega' \leqslant_a \hat{\omega} \leqslant_a \omega''$. Using this, it is easily shown that paths of minimum cost from ω_{bb} to ω_{aa}, and from ω_{aa} to ω_{bb} are given by

$$\begin{aligned}
h^\star(\omega_{bb}, \omega_{aa}) &= h^\star(\omega_{bb}, \omega_{ab}) \cup h^\star(\omega_{ab}, \omega_{aa}) \\
h^\star(\omega_{aa}, \omega_{bb}) &= h^\star(\omega_{aa}, \omega_{ab}) \cup h^\star(\omega_{ab}, \omega_{bb})
\end{aligned}$$

where $h^\star(\omega_{bb}, \omega_{ab}), h^\star(\omega_{ab}, \omega_{aa}), h^\star(\omega_{aa}, \omega_{ab})$, and $h^\star(\omega_{ab}, \omega_{bb})$ are as given above. That these are the only costly paths of the respective ω-trees with costs as given in equations 15-17 is clear. That the set of stochastically stable states are those with ω-tree of minimum cost is immediate by Lemma 4.

I now show (Theorem 3) that when ω_{ab} is stochastically stable under constant rate dynamics where both groups adapt at equal rates, the set of stochastically stable equilibria is independent of the specifics of the dynamics. What happens is this. First, varying rates of adjustment will never lower the cost $\tau^{\star}_{\omega_{ab}}$. Showing this is straightforward. Second, while it may lower the cost of $\tau^{\star}_{\omega_{bb}}$ or $\tau^{\star}_{\omega_{aa}}$, it will not lower the cost enough to alter selection. That is, Theorem 3 does not say that the minimum cost ω-tree of each convention is necessarily unchanged and as given by equations 15-17. Rather it just says that if ω_{ab} is ever stochastically stable for some constant rate dynamic, it will always be for any constant rate dynamic and it will have ω-tree with cost given by that in Theorem 2.

Theorem 3. *Suppose Condition 4 of Theorem 1 holds, so $E = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, and that the monotonic Group-Darwinian adjustment process is such that both groups evolve at constant rates. If $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$ when $\Psi^A \sim_d \Psi^B$, then $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$ for any constant rate adaptive process. Futhermore, states in $\Xi(\mathcal{G}, c_\Psi)$ have ω-tree of minimum cost equal to that as given in Theorem 2.*

Proof. The proof is found in Appendix C. □

This is a good time to mention a few features of the set up. Ellison (2000) introduced the notions of the *radius* and *coradius* of a recurrent class. The radius is defined as the minimum number of mutations necessary to escape the basin, while the coradius is defined as the maximum (over all states) of the minimum number of mutations necessary to reach the basin. When the radius is greater than the coradius, then the long run equilibrium belongs to this recurrent class. The result is not universally powerful since it does not apply in all cases. In the Language Game, it need not have relevance when $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, since it is possible that each rest point's coradius is larger than its radius.

Another observation is that due to the 2-dimensional nature of the state space, there need not be a connection between the size of each convention's basin and stochastic stability.[22] In fact, it is very possible that the equilibrium with the largest basin of attraction is not stochastically stable, and that the equilibrium with the smallest basin is stochastically stable. This can be particularly striking for parameters for which $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$.

[22] A similar though slightly different point is made in Section 8 of Young (1993).

The following example illustrates both these phenomena, and also provides the intuition behind Theorems 2 and 3. For completeness' sake, we choose non-generic parameters.

Example 5. Let $\mathcal{G}^3 = (10, 10, 4/5, 2/3)$.
$E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$ for these parameters. Using equations 5 - 8, get $(n_a^A, n_b^A, n_a^B, n_b^B) = (5, 16, 7, 13)$.

Figure 3 below illustrates the basins of attraction when both groups respond at equal rates. The non-genericity is on display by denoting the non-equilibrium rest point $(10, 3)$ by an X. The sizes of the basins of attraction are $|\mathcal{V}\big((10, 3)\big)| = 1$, $|\mathcal{V}(\omega_{bb})| = 28$, $|\mathcal{V}(\omega_{ab})| = 57$, and $|\mathcal{V}(\omega_{aa})| = 35$. Clearly $\mathcal{V}(\omega_{ab})$ is the largest.

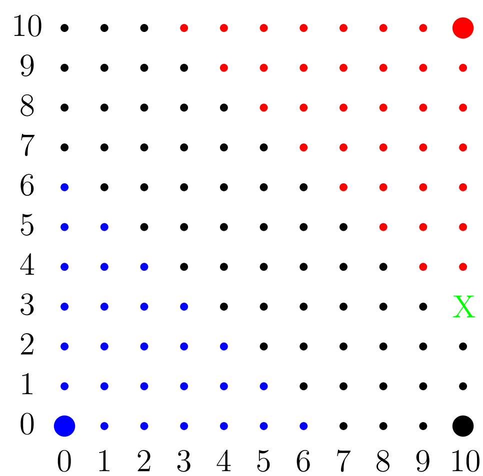

Figure 3: $\mathcal{V}(\omega_{bb}) = \bullet$. $\mathcal{V}(\omega_{ab}) = \bullet$. $\mathcal{V}(\omega_{aa}) = \bullet$.

Edges of positive cost in $\tau^\star_{\omega_{aa}}$ are $\big((0, 0) \to (7, 0)\big)$, $\big((10, 0) \to (10, 3)\big)$, and $\big((10, 3) \to (10, 4)\big)$. These have a combined cost of $7 + 3 + 1 = 11$. The costs of $\tau^\star_{(10,3)}$, $\tau^\star_{\omega_{bb}}$, and $\tau^\star_{\omega_{ab}}$ can be computed as 17, 12, and 15 respectively. By Theorem 2, ω_{aa} is the selected long run equilibrium, and this is despite it not having the maximal basin of attraction.

The [radius, coradius] pair for each absorbing state can also easily be calculated. They are given as follows, $\omega_{bb} \mapsto [7, 12]$, $\omega_{ab} \mapsto [3, 8]$, $\omega_{aa} \mapsto [7, 11]$, and $(10, 3) \mapsto [1, 10]$. Conventions ω_{bb} and ω_{aa} have equal radii, while ω_{bb} has maximal coradius. Note that all rest points have a greater coradius than radius so the Theorem of Ellison (2000) does not apply.

Let us now vary the rates of reaction for each group. First of all, note that regardless of rates, there exist some states at which the dynamics are unambiguous. These states are illustrated in Figure 4 and are colour coded by the convention to which they lead. The states not shown are a subset of $\Omega^{A,a\succ b}\cap\Omega^{B,b\succ a}$, are were part of $\mathcal{V}(\omega_{ab})$ in Figure 3 when both groups responded at equal rates. We will demonstrate how Figure 4 is modified when the groups adapt at different rates.

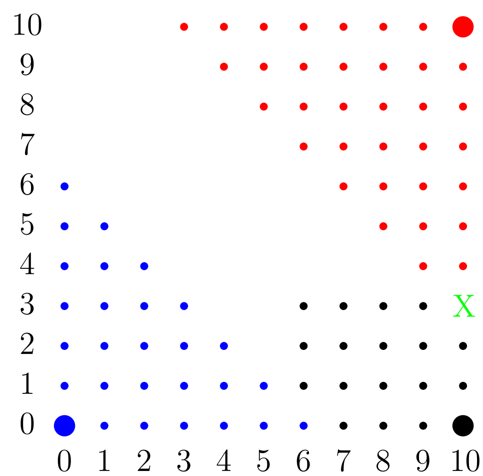

Figure 4: $\mathcal{V}(\omega_{bb}) = \bullet$. $\mathcal{V}(\omega_{ab}) = \bullet$. $\mathcal{V}(\omega_{aa}) = \bullet$.

Begin by supposing $\Psi = (\mathcal{B}^A, \Psi_1^B)$. Basins of attraction for this case are illustrated in Figure 5. Rest point $(10, 3)$ now has a non-degenerate basin of attraction, with states denoted by $\times$s. Basin $\mathcal{V}(\omega_{bb})$ is as it was when $\Psi^A \sim_d \Psi^B$. It is now the case that $\mathcal{V}(\omega_{aa}) = \Omega^{B,a\succeq b} \cup \big\{\omega \in \Omega^{A,a\succ b}\cap\Omega^{B,b\succ a} \mid [\omega]_B \geq 5\big\}$. The sizes of the basins have changed. Now, $|\mathcal{V}\big((10,3)\big)| = 7$, $|\mathcal{V}(\omega_{bb})| = 28$, $|\mathcal{V}(\omega_{ab})| = 21$, and $|\mathcal{V}(\omega_{aa})| = 65$. By Theorem 3, it must still be that $\Xi(\mathcal{G}, c_\Psi) = \{\omega_{aa}\}$. It is easily calculated that minimum cost ω-trees for the absorbing states have not changed.

While radii are always unaffected by varying reaction rates, the coradii of absorbing states ω_{aa} and $(10, 3)$ have changed. The [radius, coradius] pairs are now given as follows, $\omega_{bb} \mapsto [7, 12]$, $\omega_{ab} \mapsto [3, 8]$, $\omega_{aa} \mapsto [7, 7]$, and $(10, 3) \mapsto [1, 7]$. It is still the case that no rest point has a greater coradius than radius so the result of Ellison (2000) remains inapplicable.

Now suppose $\Psi = (\Psi_1^A, \mathcal{B}^B)$. Basins of attraction are illustrated in Figure 6. We now have $|\mathcal{V}\big((10,3)\big)| = 1$, $|\mathcal{V}(\omega_{bb})| = 61$, $|\mathcal{V}(\omega_{ab})| = 24$, and $|\mathcal{V}(\omega_{aa})| = 35$. By

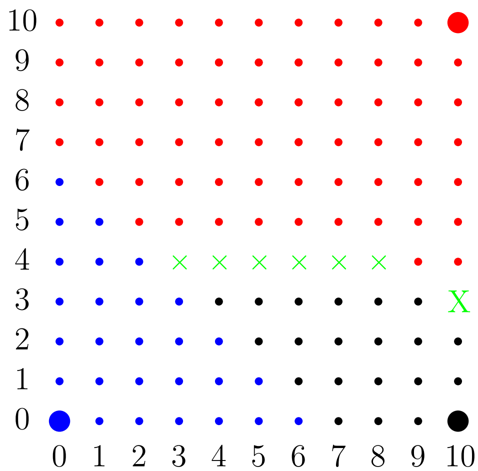

Figure 5: $\mathcal{V}(\omega_{bb}) = \bullet$. $\mathcal{V}(\omega_{ab}) = \bullet$. $\mathcal{V}(\omega_{aa}) = \bullet$.

Theorem 6, the minimum cost ω_{aa}-, ω_{ab}-, and $(10, 3)$-trees still have cost of 11, 15, and 17 respectively. Consider the minimum cost ω_{bb}-tree. The cost of 12 remains an upper bound. But an ω_{bb}-tree with cost 12 is also attainable using costly paths $\big((10, 0) \to (10, 3)\big)$, $\big((10, 3) \to (10, 4)\big)$, and $\big((10, 10) \to (2, 10)\big)$. This ω_{bb}-tree also has a total cost of 12.

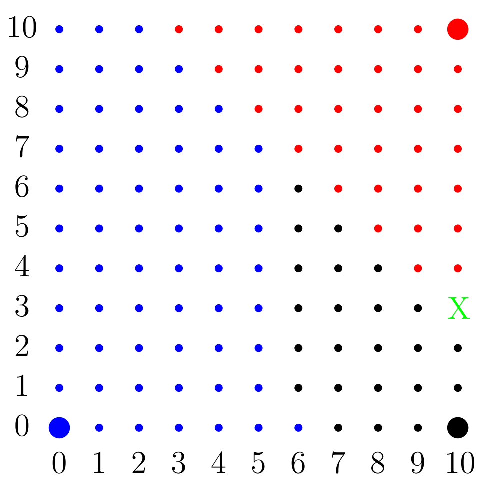

Figure 6: $\mathcal{V}(\omega_{bb}) = \bullet$. $\mathcal{V}(\omega_{ab}) = \bullet$. $\mathcal{V}(\omega_{aa}) = \bullet$.

The following modification hints at how rates can affect minimum cost ω-trees enough to alter selection. Suppose for a moment that the Group B payoffs were perturbed so that the game became generic with $(10,3) \in \Omega^{A,b\succ a}$. With $k_A = k_B$, we get $c_\Psi(\tau^\star_{\omega_{bb}}) = 12$, $c_\Psi(\tau^\star_{\omega_{ab}}) = 15$, and $c_\Psi(\tau^\star_{\omega_{aa}}) = 10$. Suppose now that $\Psi = (\Psi_1^A, \mathcal{B}^B)$. The minimum cost ω_{aa}-tree still has a cost 10. The cost of the minimum cost ω_{bb}-tree however, has lowered to 11 by proceeding along the route described in the previous paragraph. But lowering it from 12 to 11 is not enough to affect equilibrium selection in this case. However, this feature is not robust as was stated in Theorem 3.

5.2 Equilibrium set is $\{\omega_{\mathbf{bb}}, \omega_{\mathbf{aa}}\}$

Now we examine the case when only the symmetric profiles are equilibria.

Theorem 4. *Suppose Condition 4 of Theorem 1 does not hold, so $E = \{\omega_{bb}, \omega_{aa}\}$, and that the monotonic Group-Darwinian adjustment process is such that both groups evolve at constant rates, ξ_A and ξ_B respectively. Let $\tau^\star_{\omega_{bb}}$ and $\tau^\star_{\omega_{aa}}$, denote the minimum cost ω-trees for ω_{bb} and ω_{aa} respectively. Then,*

1. *If $(N^A, 0) \in \mathcal{V}(\omega_{aa})$, then*

$$
\begin{aligned}
c_\Psi(\tau^\star_{\omega_{aa}}) &= n_a^A \\
c_\Psi(\tau^\star_{\omega_{bb}}) &= n_b^A - 1_{\{\Psi^B \succ_d \Psi^A\}}\left[\max_{\omega'' \in (\mathcal{V}(\omega_{bb}))_+} \min_{\omega' \in (\Omega^{B,a\succeq b})_+} \|\omega', \omega''\|\right] \\
&= n_b^A - \left\| \left((\mathcal{V}(\omega_{bb}))_+\right)_{NW}, \left((\Omega^{A,b\succeq a})_+\right)_{NW} \right\|
\end{aligned}
$$

2. *If $(N^A, 0) \in \mathcal{V}(\omega_{bb})$, then*

$$
\begin{aligned}
c_\Psi(\tau^\star_{\omega_{bb}}) &= n_b^B \\
c_\Psi(\tau^\star_{\omega_{aa}}) &= n_a^B - 1_{\{\Psi^A \succ_d \Psi^B\}}\left[\max_{\omega'' \in (\mathcal{V}(\omega_{aa}))_-} \min_{\omega' \in (\Omega^{A,b\succeq a})_-} \|\omega', \omega''\|\right] \\
&= n_a^B - \left\| \left((\mathcal{V}(\omega_{aa}))_-\right)_{NW}, (\Omega^{A,b\succeq a})_- \right\|
\end{aligned}
$$

The set of stochastically stable states are those with the ω-tree of minimum cost. That is,

$$
\Xi(\mathcal{G}, c_\Psi) = \operatorname*{argmin}_{\omega \in \{\omega_{bb}, \omega_{aa}\}} c_\Psi(\tau^\star_\omega)
$$

Proof. The proof is found in Appendix C. □

Some things are worth noting. Clearly the state ω_{ab} plays an important role. Since it is not an equilibrium, it always lies in the basin of attraction of the larger group's preferred convention. If $N^A > N^B$ and $\Psi^A \succ_d \Psi^B$, the set $\Omega \backslash (\Omega^{A,b \succeq a} \cup \Omega^{B,a \succeq b})$ is a subset of $\mathcal{V}(\omega_{aa})$. When $\omega_{ab} \in \mathcal{V}(\omega_{aa})$, then the selected long run equilibrium might change if $\Psi^B \succ_d \Psi^A$. When $\omega_{ab} \in \mathcal{V}(\omega_{bb})$ the reverse holds. Appendix B shows how to construct basins of attraction when the smaller group responds at a faster rate.

Suppose Group A is the larger group. The minimum cost of an ω_{aa}-tree is n_a^A (since $(n_a^A, 0) \in \mathcal{V}(\omega_{aa})$), while a lower bound for the minimum cost ω_{bb}-tree is n_b^B (if $(N^A - n_b^B, N^B)$). Thus, Group B must have a stronger relative preference for its preferred equilibrium to reverse the long run outcome. Conversely, if Group A has an equally strong preference, then $(\mathbf{a}, \mathbf{a})$ will always be a stochastically stable state.

Theorem 4 has a nice geometric interpretation, that can be seen by referring back to Figures 1 and 2 which regard Example 1. From equation 5, $n_a^A = 7 \leq N^A$, and therefore $(N^A, 0) \in \mathcal{V}(\omega_{aa})$. Figure 1 represents the basins of attraction when $k_A = k_B$. In this case, the minimum cost ω_{aa}-tree has cost of 7 - the costly edge of the path being $\big(\omega_{bb} \to (7,0)\big)$. Similarly the minimum cost ω_{bb}-tree has cost of 9 - the costly edge of the path being $\big(\omega_{aa}, (1,5)\big)$. Thus when $k_A = k_B$, clearly ω_{aa} is the stochastically stable outcome.

Now look at Figure 2 representing the basins of attraction when $\Psi = (\Psi_1^A, \Psi_2^B)$. The minimum cost ω_{aa}-tree is unchanged, since the transition $\big(\omega_{bb} \to (7,0)\big)$, is still an edge on a path of minimum cost from ω_{bb} to $\mathcal{V}(\omega_{aa})$. The cost of the minimum cost ω_{bb}-tree is different to when $k_A = k_B$, since now the transition $\big(\omega_{aa} \to (3,5)\big)$ is the first edge on a path of minimum cost from ω_{aa} to $\mathcal{V}(\omega_{bb})$. This has a cost of 7. So the long run distribution assigns positive probability to both ω_{bb} and ω_{aa}.

Consider Figure 1 and suppose that $\Psi = (\Psi_1^A, \Psi_3^B)$. Note that $\omega_{ba} \in \Omega^{A,b \succ a}$ and so we use the second method described above. The example is not generic and $(N^A, N^B) = (10, 5)$, so that

$$\big((k_A, r_A), (k_B, r_B)\big) = \big((5,0), (1,2)\big)$$

Clearly then, $\hat{k} = 1$, and $\Psi^{\leftarrow \hat{k}}\big((B^0)_+\big) = \Psi^{\leftarrow \hat{k}}\big((6,0)\big) = (5,3)$, and by equation 20,

$(B^\infty)_{NW} = (B^{\hat{n}})_{NW} = (B^2)_{NW} = (4,5)$. Describing the sets we have that

$$
\begin{aligned}
B^0 &= \Omega^{A,b\succeq a} \\
B^1 &= B^0 \cup \{\omega \mid [\omega]_A + [\omega]_B = 8, [\omega]_A \le 5\}^{\downarrow} \\
B^2 &= B^1 \cup \{(4,5)\}^{\downarrow} \\
B^3 &= B^2
\end{aligned}
$$

Now, we have that $\mathcal{V}(\omega_{bb}) = \Omega^{A,b\succeq a} \cup (B^{\hat{n}} \cap \Omega^{A,a\succ b}) = \{(3,5),(4,4),(5,3),(6,1)\}^{\downarrow}$.

Having classified conditions for which the various equilibria are selected, the next Section compares the selected equilibrium to those a planner might want to induce.

6 Stability versus Welfare

When the population is homogeneous, players collectively agree on what action they "would like to have taken" earlier today, and hence what they will choose for tomorrow if afforded a revision opportunity. The main issue in existing large population coordination problems is the tension between efficiency and risk-dominance. Both KMR and Young (1993) show that the risk dominant action will emerge under perturbed best response based dynamics. This result is *negative* in the sense that the locally risk-dominant equilibrium action need not coincide with the pareto dominant one, and may have payoffs that are dominated to an arbitrary extent.

In the Language Game, pareto efficiency is useless as a selection device. Both symmetric profiles are socially efficient equilibria, and there is never uniform preference over these. The purpose of this Section is to rank profiles in $E(\mathcal{G})$ according to various welfare criteria, and then compare this ranking to the outcome(s) selected by the dynamics. To infer how members of the population rank profiles in $E(\mathcal{G})$, it suffices to analyse the situation from the perspective of any one agent from each Group.

Theorem 5. *Within the set of group-symmetric profiles:*[23]

1. *ω_{aa} and ω_{bb} are always socially efficient.*

[23] Again, this Theorem is stated for correctly specified utility functions, and not those as given by equations (1)-(4). See footnote **??**.

2. ω_{ab} is socially efficient if and only if

$$p \geq \frac{N-1}{2N-N^B-2} \quad \textit{and} \quad q \geq \frac{N-1}{2N-N^A-2} \tag{18}$$

Proof. The proof is straightforward and is omitted. □

A natural question to ask is the relationship between social efficiency and equilibrium. It turns out that a socially efficient profile must be an equilibrium, but not all equilibria are socially efficient.

Theorem 6. *If ω_{ab} is socially efficient, then it must be an equilibrium. But ω_{ab} may be an equilibrium without being socially efficient.*

Proof. Follows from conditions in part 4 of Theorem 1, and those in part 2 of Theorem 5. Clearly the second implies the first, while the first need not imply the second. □

Recall the examples from Section 3. In Example 1, with $\mathcal{G}^1 = (10, 5, 3/5, 3/5)$ and $E(\mathcal{G}^1) = \{\omega_{bb}, \omega_{aa}\}$, we have $\{U^A(\omega_{aa}), U^B(\omega_{bb})\} = \{42/5, 28/5\}$, and $\{U^A(\omega_{bb}), U^B(\omega_{bb})\}$ $= \{28/5, 42/5\}$. The 10 Group A members desire ω_{aa}, while the 5 in Group B prefer ω_{bb}. In Example 2, with $\mathcal{G}^2 = (10, 5, 3/5, 5/6)$ and $E(\mathcal{G}^2) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, we get $\{U^A(\omega_{aa}), U^B(\omega_{aa})\} = \{42/5, 28/6\}$, $\{U^A(\omega_{ab}), U^B(\omega_{ab})\} = \{27/5, 50/6\}$, and $\{U^A(\omega_{bb}), U^B(\omega_{bb})\} = \{28/5, 140/6\}$. Note that ω_{ab} is pareto dominated by ω_{bb}, and so provides an example of Theorem 6 at work.

The next obvious question is whether or not a decentralized adjustment process can ever select a socially-inefficient convention. In Example 2, ω_{ab} is socially inefficient but not stochastically stable. Theorem 7 shows that this is generalizable.

Theorem 7. *Suppose $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$. For any Group-Darwinian adjustment process, Ψ, it will never be the case that ω_{ab} is socially inefficient and yet $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$.*

Proof. The proof is found in Appendix C. □

There are a host of other commonly used social welfare criteria from the theory of social choice. Other interesting ones to apply here are the Utilitarian Welfare function, and the Rawlsian Welfare function. Under these criteria, inefficient outcomes can certainly emerge.

7 Comparative Statics[24]

A few obvious questions spring to mind. Would a player always prefer to be part of a larger group? Would a player always prefer to be part of a group with stronger relative preference?

7.1 Varying Payoffs

The full treatment of this issue is left to future research. In this subsection I provide only an illustrative example.

Example 6 (Increased Payoffs)**.** Let $\mathcal{G}^4 = (10, 10, 4/5, 3/5)$, and Ψ be such that $\Psi^A \sim_d \Psi^B$.

Using equations (5)-(8) it can be calculated that $(n_a^A, n_b^A, n_a^B, n_b^B) = (5, 12, 16, 9)$, and so by Theorem 2, it must be that $\Xi(\mathcal{G}^4, c_\Psi) = \{\omega_{aa}\}$. Clearly, average payoffs at this point are given by $U^A(\omega_{aa}) = 4/5$ and $U^B(\omega_{bb}) = 2/5$.

Now suppose that Group B payoffs are modified such that their preference for action b is magnified. Precisely, suppose that $\mathcal{G}^4$ is transformed to $\hat{\mathcal{G}}^4 = (10, 10, 4/5, 4/5)$. It can now be calculated that $(n_a^A, n_b^A, n_a^B, n_b^B) = (5, 16, 16, 5)$, and so by Theorem 2, it is now the case that $\Xi(\hat{\mathcal{G}}^4, c_\Psi) = \{\omega_{ab}\}$. Average payoffs now are given by $U^A(\omega_{ab}) = U^B(\omega_{ab}) = (9/19)(4/5) < 2/5$. And so Group B's payoffs have gone down.

Now suppose that the Group B payoffs are modified again, with their preference for action b increased further still. Precisely, suppose that $\hat{\mathcal{G}}^4$ is transformed to $\hat{\hat{\mathcal{G}}}^4 = (10, 10, 4/5, 19/20)$. Now, $(n_a^A, n_b^A, n_a^B, n_b^B) = (5, 16, 19, 2)$, and by Theorem 2, it is still the case that $\Xi(\hat{\hat{\mathcal{G}}}^4, c_\Psi) = \{\omega_{ab}\}$. Average payoffs now are given by $U^A(\omega_{ab}) = (9/19)(4/5)$ and $U^B(\omega_{ab}) = 9/20 > 2/5$. Thus Group B's average payoff has increased.

7.2 Varying Group Size

[24] [This section is incomplete and will be finished soon]

8 Conclusion

This paper introduces a new coordination game in the hope of shedding some light on how behaviour might develop in societies with heterogeneous agents. The environment, the "Language Game", deviates from existing large population models in one simple but important way: there are two distinct homogeneous groups with pairwise interactions occurring both within- and across-group.

Three properties matter for equilibrium selection in the Language Game. They are (i) group size, (ii) group payoffs, and (iii) the rates at which the groups react - "group dynamics". Any agent always desires to be part of a more reactive group, but does not always long for greater numbers in their group or for more polarized preferences. While interesting, the results are not robust. That is, for a given game, assumptions on the likelihood of behavioural mutations, and the full connectedness of the population were essential.

However, the fragility of the results may not be the weakness it first appears. For example, while the results of Bergin and Lipman (1996) show that any equilibrium can be selected for appropriately defined mutations, I know of no examples where the Ising model dynamics of Blume (1993) select *different* long run equilibria to those of KMR. But they can be shown to for an open set of parameters in the Language Game (Neary, 2010a). Another major limitation of existing large population pure coordination problems, is that equilibrium selection is robust to network architecture, with uniform adoption of the locally risk-dominant action stochastically stable for any network (Peski, 2010). An immediate consequence of moving the Language Game to arbitrary networks, is that both network topology and the specifics of the dynamics matter strongly for equilibrium selection (Neary, 2010a).

Evolutionary game theory has typically focused on the many different and interesting ways in which behaviour adapts in large populations. However, in my opinion, there has been too little attention on whether or not the stage games accurately capture all situations in which large populations engage. In a companion paper, (Neary, 2010b), I define a new class of large population games called "Multiple-Group Games" (MGGs). The key feature of a MGG is that the population is partitioned into groups, with players interacting pairwise with potentially anyone from the population. The only constraint is that within-group interactions must be symmetric. Across-group interactions can be anything, as long as each player has the same strategy set available in each. This adds heterogeneity in a surprisingly tractable way. Perhaps the greatest advantage of this framework is the number of new possibilities it introduces and the number of extensions it permits.

APPENDIX

A Lattices

Let ω and ω' be elements of Ω. Write $\omega \geqslant_a \omega'$ if $[\omega]_A \geq [\omega']_A$ and $[\omega]_B \geq [\omega']_B$, and $\omega \not\geqslant_a \omega'$ if $[\omega]_A < [\omega']_A$ or $[\omega]_B < [\omega']_B$. Write $\omega >_a \omega'$ if $\omega \geqslant_a \omega'$ and $\omega \neq \omega'$, and $\omega \gg_a \omega'$ if both $[\omega]_A > [\omega']_A$ and $[\omega]_B > [\omega']_B$. States ω' and ω'' are *comparable*, $\omega'' \perp_a \omega'$, if $\omega' \geqslant_a \omega''$ or $\omega'' \geqslant_a \omega'$ or both, while ω' and ω'' are *incomparable*, $\omega' \parallel_a \omega''$, if $\omega' \not\geqslant_a \omega''$ and $\omega'' \not\geqslant_a \omega'$. The pair $(\Omega, \geqslant_a)$ is a complete lattice with *bottom* element ω_{bb} and *top* element ω_{aa}.

A nonempty $\Lambda \subset \Omega$ is a *chain*, if for all $\omega', \omega'' \in \Lambda$, $\omega' \perp_a \omega''$. For a given chain Λ, define $(\Lambda)_{NE} = \{\omega \in \Lambda \,|\, \forall \omega' \in \Lambda,\, \omega \geqslant_a \omega'\}$, and $(\Lambda)_{SW} = \{\omega \in \Lambda \,|\, \forall \omega' \in \Lambda,\, \omega \leqslant_a \omega'\}$. A nonempty $\Upsilon \subset \Omega$ is an *anti-chain*, if for any $\omega', \omega'' \in \Upsilon$ with $\omega' \neq \omega''$, it is the case that $\omega' \parallel_a \omega''$. For a given anti-chain Υ, define $(\Upsilon)_{SE} = \{\omega \in \Upsilon \,|\, \forall \omega' \in \Upsilon,\, [\omega]_A \geq [\omega']_A\}$, and $(\Upsilon)_{NW} = \{\omega \in \Upsilon \,|\, \forall \omega' \in \Upsilon,\, [\omega]_B \geq [\omega']_B\}$.

A *down set* is a nonempty set $D \subseteq \Omega$, where if $\omega' \in D$, and $\omega'' \in \Omega$ is such that $\omega'' \leqslant_a \omega'$, then $\omega'' \in D$. A *principal down set* is a down set of the form $\{\omega\}^{\downarrow} := \{\omega' \in \Omega \,|\, \omega' \leqslant_a \omega\}$. Similarly, an *up set* is a nonempty set $U \subseteq \Omega$, where if $\omega' \in U$ and $\omega'' \in \Omega$ is such that $\omega'' \geqslant_a \omega'$, then $\omega'' \in U$. A *principal up set* is an up set $\{\omega\}^{\uparrow} := \{\omega' \in \Omega \,|\, \omega' \geqslant_a \omega\}$. The intersection of two principal down sets is a principal down set, and $\omega' \wedge \omega''$ denotes the element, referred to as the *meet* of ω' and ω'', such that $\{\omega'\}^{\downarrow} \cap \{\omega''\}^{\downarrow} = \{\omega' \wedge \omega''\}^{\downarrow}$. Similarly, $\omega' \vee \omega''$, referred to as the *join* of ω' and ω'', denotes the element where $\{\omega'\}^{\uparrow} \cap \{\omega''\}^{\uparrow} = \{\omega' \vee \omega''\}^{\uparrow}$. For any set $Q \subseteq \Omega$, use $Q^{\downarrow}$ ($Q^{\uparrow}$) to denote the down (up) set generated by Q. That is, $Q^{\downarrow} = \cup_{\omega \in Q} \{\omega\}^{\downarrow}$.

If $\omega', \omega'' \in \Omega$ are such that $\omega' \leqslant_a \omega''$, then the *interval* $[\omega', \omega'']$ is defined as $[\omega', \omega''] := \{\omega'\}^{\uparrow} \cap \{\omega''\}^{\downarrow} = \{\omega \in \Omega \,|\, \omega' \leqslant_a \omega \leqslant_a \omega''\}$. If $\omega' \parallel \omega''$, then $[\omega', \omega''] = \emptyset$. A nonempty subset $Q \subseteq \Omega$ is said to be *convex* if $[\omega', \omega''] \subseteq Q$ for all $\omega', \omega'' \in Q$ with $\omega' \leqslant_a \omega''$. Both up sets and down sets are convex.

For any nonempty $Q \subseteq \Omega$, define the *lower boundary* of Q by $(Q)_- := \{\omega \in Q \,|\, \{\omega\}^{\downarrow} \cap Q = \omega\}$. Similarly define the *upper boundary* by $(Q)_+ := \{\omega \in Q \,|\, \{\omega\}^{\uparrow} \cap Q = \omega\}$. The lower and upper boundaries each form an anti-chain. For any pair $\omega', \omega'' \in Q^{\downarrow}$, say that $\omega \in Q$ is *join-irreducible* in Q, if $\omega' \vee \omega'' = \omega$ implies $\omega' = \omega$ or $\omega'' = \omega$. Similarly, for any pair $\omega', \omega'' \in Q^{\uparrow}$, say that $\omega \in Q$ is *meet-irreducible* in Q, if $\omega' \wedge \omega'' = \omega$ implies $\omega' = \omega$ or $\omega'' = \omega$. The sets of join-irreducible and meet-irreducible elements of Q are denoted $(Q)_{--}$ and $(Q)_{++}$, and referred to as the *total lower boundary* and the *total upper boundary* of Q respectively. $(Q)_{--}$ and $(Q)_{++}$ can each be viewed as a union of "horizontal" row chains $r_1, \ldots, r_m$ or as a union of "vertical" column chains $c_1, \ldots, c_n$, where for any $\omega_{i_k} \in c_i$ and $\omega_{j_l} \in c_j \neq c_i$, $\omega_{i_k} \parallel \omega_{j_l}$.

Define a metric, $\|\| : \Omega \times \Omega \to \{0, \ldots, N\}$, on Ω as follows. For any $\omega', \omega'' \in \Omega$, let $\|\omega', \omega''\| = \sum_{K=A,B} |\,[\omega']_K - [\omega'']_K\,|$. For a given set $Q \subseteq \Omega$ and state $\omega \notin Q$, interest will often focus on the state(s) in Q, that is (are) closest to ω. That is, in the set $\left\{\omega' \in Q \,|\, \omega' \in \operatorname{argmin}_{\hat{\omega} \in Q} \|\omega, \hat{\omega}\|\right\}$. Most importantly, for a given convex set Q, and an element $\omega \notin Q$ such that there exists an element $\omega' \in (Q)_{++}$ with $\omega \geqslant_a \omega'$, then $\min_{\hat{\omega} \in Q} \|\omega, \hat{\omega}\|$ is attained by some $\omega^\star \in (Q)_{++}$.

B Constructing Basins of Attraction

Let 2^Ω denote the power set of Ω. Given a mapping $\Psi : \Omega \to \Omega$, consider the *inverse image map* $\Psi^{\leftarrow} : 2^\Omega \to 2^\Omega$ defined for every nonempty $Q \subseteq \Omega$ by $\Psi^{\leftarrow}(Q) = \{\omega \,|\, \Psi(\omega) \in Q\}$. In constructing the basins of attraction, the following result is used repeatedly.

Lemma 6. *The following are equivalent:*

1. *If $L \subseteq \Omega$ is a down set, then $\Psi^{\leftarrow}(L)$ is a down set*

2. *Ψ is monotone.*

Proof. $(1 \Rightarrow 2)$ It is enough to restrict attention to principal lower sets, since if $(\{\omega_1\}^{\downarrow}, \ldots, \{\omega_n\}^{\downarrow})$ is a collection of principal lower sets, then $\cup_{j=1}^{n} \{\omega_j\}^{\downarrow}$ is a down set. For every $\omega \in \Omega$ we have that $\omega \in \Psi^{\leftarrow}\big(\{\Psi(\omega)\}^{\downarrow}\big)$. This by assumption is a lower set, so if $\omega' \in \Omega$ is such that $\omega' \leqslant_a \omega$, then $\omega' \in \Psi^{\leftarrow}\big(\{\Psi(\omega)\}^{\downarrow}\big)$. Clearly $\Psi(\omega') \leqslant_a \Psi(\omega)$ and hence Ψ is monotone.

$(2 \Rightarrow 1)$ Let $\omega \in \Omega$ and define $Q = \Psi^{\leftarrow}\big(\{\Psi(\omega)\}^{\downarrow}\big)$. Provided $Q \neq \emptyset$, consider some $\omega' \in Q$. Clearly for every $\omega'' \in \Omega$ with $\omega'' \leqslant_a \omega'$, we have that $\Psi(\omega'') \leqslant_a \Psi(\omega') \leqslant_a \omega$, which gives that $\omega'' \in Q$. Thus Q is a down set. Finally, we note that for any $L \subseteq \Omega$, it is the case that $\Psi(L) = \Psi(\cup_{\omega' \in L} \{\omega'\}) = \cup_{\omega' \in L} \Psi(\omega')$, and combining this with the fact that the union of a collection of lower sets is a lower set yields the desired result.

□

If neither condition of Lemma 2 is satisfied, then Ω can be partitioned as $\big\{\Omega^{A,b\succ a}, \Omega^{A,a\succ b}\big\} = \big\{\Omega^{B,b\succ a}, \Omega^{B,a\succ b}\big\}$. Different monotonic Group-Darwinian adjustment processes have no affect on basins of attraction in this case. Thus both partitions are equivalent to $\{\mathcal{V}(\omega_{bb}), \mathcal{V}(\omega_{aa})\}$.

So suppose one of the sufficient conditions of Lemma 2 holds, and consider the partition of Ω $\big\{\Omega^{A,b\succeq a}, \Omega^{B,a\succeq b}, \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a})\big\}$. I show how to construct $\mathcal{V}(\omega_{bb})$ (the construction of $\mathcal{V}(\omega_{aa})$ follows analogously). There is no need to show how to construct $\mathcal{V}(\omega_{ab})$ when $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, since $\mathcal{V}(\omega_{ab}) = \Omega \backslash \big(\cup_{\omega \in \Omega_0 \backslash \{\omega_{ab}\}} \mathcal{V}(\omega)\big)$. Begin by constructing the following sequence of states:

$$B^0 := \begin{cases} \Omega^{A,b\succeq a}, & \text{genA} \\ \Omega^{A,b\succeq a} \backslash \big\{(n_a^A - 1, 0)\big\}, & \text{ngenA} \end{cases}$$

and for each $n \geq 1$, define $B^n := \Psi^{\leftarrow}\big(B^{n-1}\big)$. Finally, define $B^{\infty} := \bigcup_{n=0}^{\infty} B^n$.

Lemma 7. *For any monotonic Group-Darwinian dynamic Ψ,*

1. *For all $n \geq 0$, $B^n \subseteq B^{n+1}$.*

2. *$\Psi^{\leftarrow}\big(B^{\infty}\big) = B^{\infty} = \mathcal{V}(\omega_{bb})$.*

Remark: By part 1, the sets $\{B^n\}$ are weakly increasing in n. By part 2 the iterative procedure eventually it stops. That is, there exists an $\hat{n} \in \mathbb{N}$ such that for all $n \geq \hat{n}$, $B^n = B^{\hat{n}}$.

Proof. 1. The proof is by induction. It is true by definition for $n = 0$. By Lemma 6 and the fact that $\Psi(\omega) \leqslant_a \omega$ for each $\omega \in \Omega^{A,b\succeq a}$, it must be that for any lower set $Q \subseteq \Omega^{A,b\succeq a}$, that $Q \subseteq \Psi^{\leftarrow}(Q)$. And so the claim is true for $n = 1$. Assume it is true for $n = k$. We have

$$\begin{aligned} B^{k+2} &= \Psi^{\leftarrow}\big(B^{k+1}\big) \\ &= \Psi^{\leftarrow}\big(B^k \cup (B^{k+1} \backslash B^k)\big) \\ &= \Psi^{\leftarrow}\big(B^k\big) \cup \Psi^{\leftarrow}\big(B^{k+1} \backslash B^k\big) \\ &\supseteq B^{k+1} \end{aligned}$$

where the third equality follows from the definition of $\Psi^{\leftarrow}$, and the inclusion from the inductive step. Thus the claim holds for $k + 1$.

2. The first equality is clear using part 1 and the fact that Ω is finite. The second inequality follows due to the convexity of the basins of attraction (Lemma 3).

$\square$

The entire class of monotonic Group-Darwinian adjustment processes is too large to manage in a tractable way, and so I restrict attention to those that vary at constant rates. Any constant rate dynamic, Ψ, satisfies one of two properties: $\Psi^A \succeq_d \Psi^B$, or $\Psi^B \succ_d \Psi^A$. I consider each case separately.

There are three types of scenario to consider: those with $E(\mathcal{G}) = \{\omega_{bb}, \omega_{aa}\}$ and $\omega_{ab} \in \Omega^{B,a\succ b}$; those with $E(\mathcal{G}) = \{\omega_{bb}, \omega_{aa}\}$ but instead $\omega_{ab} \in \Omega^{A,b\succ a}$; and those where $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$. I ignore the second of these for reasons that will become clear soon.

1. $\Psi^A \succeq_d \Psi^B$.
   When $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, or $E(\mathcal{G}) = \{\omega_{bb}, \omega_{aa}\}$ with $\omega_{ab} \in \Omega^{B,a\succ b}$, it is clear that for any $\omega \in \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a}$, we have $\Psi(\omega) \in \Omega^{A,a\succ b}$. And so by Lemma 9 it must be that, $\mathcal{V}(\omega_{bb}) = D^\infty = B^\infty = B^0$

2. $\Psi^B \succ_d \Psi^A$.

   Since for any $\omega \in \left(\left\{(n_a^A - 1, 0)\right\}^\uparrow \setminus \left\{(n_a^A - 1, 0)\right\}\right)$, it must be that $[\Psi(\omega)]_A \geq n_a^A$, we have

$$\mathcal{V}(\omega_{bb}) \subseteq \begin{cases} \left\{(n_a^A - 2, N^B)\right\}^\downarrow \cup (n_a^A - 1, 0), & \text{genA} \\ \left\{(n_a^A - 3, N^B)\right\}^\downarrow \cup \left\{(n_a^A - 2, 1)\right\}^\downarrow, & \text{ngenA} \end{cases}$$

   The goal is to calculate $\left(\mathcal{V}(\omega_{bb})\right)_{++}$. To this end, I proceed in a slightly roundabout way. First it will be useful to define the following operator, $\Phi = (\Phi^A_{\xi_A}, \Phi^B_{\xi_B}) : \Omega \to \mathbb{N} \times \mathbb{N}$, where for each $K \in \{A, B\}$

$$\Phi^K(\omega) = \begin{cases} [\omega]_K - \xi_K, & \text{if } U^K(a;\omega) < U^K(b;\omega), \\ [\omega]_K, & \text{if } U^K(a;\omega) = U^K(b;\omega), \\ [\omega]_K + \xi_K, & \text{if } U^K(a;\omega) > U^K(b;\omega), \end{cases}$$

   Clearly, $\Phi = (\Phi^A_{\xi_A}, \Phi^B_{\xi_B})$ agrees with constant rate process $\Psi = (\Psi_{\xi_A}, \Psi^B_{\xi_B})$ when the range of Φ is restricted to $\left\{0, \ldots, N^A\right\} \times \left\{0, \ldots, N^B\right\}$. Now define $D^0 = B^0$ and let us perform repeated applications of $\Phi^\leftarrow$ on D^0. That is, for all $n \leq \hat{n}$ define $D^n = \Phi^\leftarrow(D^{n-1}) \cap \Omega$. To help manage this, define $C^0 = (D^0)_+$, and for each $n < \hat{n}$, define the following $C^{n+1} := (D^{n+1})_+ \setminus (D^n)_+$. That C^n is a nonempty anti-chain for each $n < \hat{n}$ is clear. It should also be clear that for all $n < \hat{n} - 1$, $(D^{n+1})_+ = (\cup_{j=0}^n D^j)_+$.

   Now we are ready to proceed in calculating $D^{\hat{n}}$. Recall that Groups A and B respond at constant rates ξ_A and ξ_B respectively. There are two cases to consider

   (a) $\omega_{ba} \in \Omega^{A,a\succ b}$.
   Define (k_A, k_B) as follows:

$$(k_A, k_B) = \begin{cases} \left(\left\lfloor \frac{n_a^A - 1}{\xi_A} \right\rfloor, \left\lfloor \frac{N^B}{\xi_B} \right\rfloor\right), & \text{genA} \\ \\ \left(\left\lfloor \frac{n_a^A - 2}{\xi_A} \right\rfloor, \left\lfloor \frac{N^B - 1}{\xi_B} \right\rfloor\right), & \text{ngenA} \end{cases}$$

Now define $\hat{k} = \min\{k_A, k_B\}$ and note that when $\hat{k} \geq 1$, then for all $k \leq \hat{k}$

$$\Phi^{\leftarrow^k}\Big((C^0)_{SE}\Big) = \begin{cases} \Big(n_a^A - 1 - k\xi_A, k\xi_B\Big), & \text{genA} \\ \\ \Big(n_a^A - 2 - k\xi_A, k\xi_B + 1\Big), & \text{ngenA} \end{cases}$$

This element $(C^0)_{SE}$ is the key element, since for all $n \leq \hat{k}-1$, $(C^{n+1})_{SE} = \Psi^{\leftarrow}\big(\{(C^n)_{SE}\}\big)$. For all $n \leq \hat{k} = \hat{n} - 1$, $D^n = D^{n-1} \cup \Phi^{\leftarrow}(C^{n-1})^{\downarrow}$. Difficulties arise in calculating $C^{\hat{n}}$ and hence $D^{\hat{n}}$, since it must be that $\Psi^{\leftarrow}((C^{\hat{n}-1})_{SE}) = \emptyset$. Precisely, these difficulties arise at the state $\Phi^{\leftarrow}\big(\Phi^{\leftarrow^{\hat{k}}}((C^0)_+)\big)$, since by definition one of the following must be the case

$$\begin{aligned} &\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_A - \xi_A < 0, \text{ or} \\ &\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_B + \xi_B > N^B \end{aligned}$$

So there are 3 cases to consider: $k_B < k_A$, $k_B > k_A$, and $k_B = k_A$. Before analysing each case, define the "remainders", r_A and r_B as follows

$$(r_A, r_B) = \begin{cases} \Big(n_a^A - 1 - \hat{k}\xi_A, N^B - \hat{k}\xi_B\Big), & \text{genA} \\ \\ \Big(n_a^A - 2 - \hat{k}\xi_A, N^B - 1 - \hat{k}\xi_B\Big), & \text{ngenA} \end{cases} \tag{19}$$

- $k_B < k_A$.
  In this case, $\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_A - \xi_A > 0$, but $\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_B + \xi_B \geq N^B$, and so

$$\big((D^{\hat{n}})_+\big)_{NW} = \begin{cases} \Big(n_a^A - 1 - (\hat{k}+1)\xi_A, N^B\Big), & \text{genA} \\ \\ \Big(n_a^A - 2 - (\hat{k}+1)\xi_A, N^B\Big), & \text{ngenA} \end{cases}$$

- $k_B > k_A$.
  In this case, $\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_A - r_A \leq 0$, and $\Big[\Psi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_B + \xi_B < N^B$, and so

$$\big((D^{\hat{n}})_+\big)_{NW} = \begin{cases} \Big(0, \hat{k}\xi_B + r_A\Big), & \text{genA} \\ \\ \Big(0, \hat{k}\xi_B + 1 + r_A\Big), & \text{ngenA} \end{cases}$$

- $k_B = k_A$
  In this case we have that $\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_A - \xi_A < 0$, and $\Big[\Phi^{\leftarrow^{\hat{k}}}\big((C^0)_+\big)\Big]_B + \xi_B > N^B$.
  There are two cases to consider,
  - if $r_A < r_B$, then

$$\big((D^{\hat{n}})_+\big)_{NW} = \begin{cases} \Big(0, \hat{k}\xi_B + r_A\Big), & \text{genA} \\ \\ \Big(0, \hat{k}\xi_B + r_A + 1\Big), & \text{ngenA} \end{cases}$$

– if $r_A \geq r_B$, then

$$\left((D^{\hat{n}})_+\right)_{NW} = \begin{cases} \left(n_a^A - 1 - \hat{k}\xi_A - r_B, N^B\right), & \text{genA} \\ \\ \left(n_a^A - 2 - \hat{k}\xi_A - r_B, N^B\right), & \text{ngenA} \end{cases}$$

(b) $\omega_{ba} \in \Omega^{A, b \succeq a}$.
This time, define (k_A, k_B) as follows:

$$(k_A, k_B) = \begin{cases} \left(\left\lfloor \frac{N^B}{\xi_A} \right\rfloor, \left\lfloor \frac{N^B}{\xi_B} \right\rfloor\right), & \text{genA} \\ \\ \left(\left\lfloor \frac{N^B - 1}{\xi_A} \right\rfloor, \left\lfloor \frac{N^B - 1}{\xi_B} \right\rfloor\right), & \text{ngenA} \end{cases}$$

Now define $\hat{k} = \min\{k_A, k_B\}$, and again note that when $\hat{k} \geq 1$, then for all $k \leq \hat{k}$

$$\Phi^{\leftarrow k}\left((C^0)_+\right) = \begin{cases} \left(n_a^A - 1 - k\xi_A, k\xi_B\right), & \text{genA} \\ \\ \left(n_a^A - 2 - k\xi_A, k\xi_B + 1\right), & \text{ngenA} \end{cases}$$

When $\omega_{ba} \in \Omega^{A, a \succ b}$ and $\Psi^B \succ_d \Psi^A$, there are only two cases to consider: $\hat{k} = k_B < k_A$, and $\hat{k} = k_B = k_A$. Remainders are again defined as in equation 19.

- $k_B < k_A$.
  In this case, $\left[\Phi^{\leftarrow \hat{k}}\left((C^0)_+\right)\right]_A - \xi_A > n_{Aa} - 1 - N^B$, but $\left[\Phi^{\leftarrow \hat{k}}\left((C^0)_+\right)\right]_B + \xi_B \geq N^B$, and so

$$\left((D^{\hat{n}})_+\right)_{NW} = \begin{cases} \left(n_a^A - 1 - (\hat{k}+1)\xi_A, N^B\right), & \text{genA} \\ \\ \left(n_a^A - 2 - (\hat{k}+1)\xi_A, N^B\right), & \text{ngenA} \end{cases}$$

- $k_B = k_A$
  In this case we have that $\left[\Phi^{\leftarrow \hat{k}}\left((C^0)_+\right)\right]_A - \xi_A < n_{Aa} - 1 - N^B$, and $\left[\Phi^{\leftarrow \hat{k}}\left((C^0)_+\right)\right]_B + \xi_B > N^B$. However, it must be the case that $r_A \geq r_B$, in which case

$$\left((D^{\hat{n}})_+\right)_{NW} = \begin{cases} \left(n_a^A - 1 - \hat{k}\xi_A - r_B, N^B\right), & \text{genA} \\ \\ \left(n_a^A - 2 - \hat{k}\xi_A - r_B, N^B\right), & \text{ngenA} \end{cases} \tag{20}$$

It is almost immediate that $\mathcal{V}(\omega_{bb}) = B^{\hat{n}} = D^{\hat{n}} \cap \Omega^{A, a \succ b}$, and so we have computed what we set out to achieve.

C Proofs

Proof of Lemma 2.

It is enough to show the existence of a state $\omega \in \Omega$ such that $\operatorname{sign}\big(U^A(a;\omega)-U^A(b;\omega)\big) \neq \operatorname{sign}\big(U^B(a;\omega)-U^B(b;\omega)\big)$. This is implied by the stronger conditions that both $n_a^B > n_a^A$ and $n_b^A > n_b^B$. I show only the first of these. Note that

$$\begin{aligned} n_a^B &= \Big\lceil q(N-2)+1 \Big\rceil \\ &\geq \Big\lceil (1-p)(N-2)+1 \Big\rceil \\ &= n_a^A \end{aligned}$$

1. Observe that the weak inequality is strict when N is even.
2. When $n_a^B = \lceil \alpha \rceil$ and $n_a^A = \lceil \beta \rceil$, $\alpha, \beta \in \mathbb{R}$, a sufficient condition for $n_a^B > n_a^A$ is that $\alpha > \beta + 1$. This sufficient condition yields the required conclusion since

$$\begin{aligned} n_a^B > n_a^A &\iff\!\!\!\!\!\!\Leftarrow\; q(N-2)+1 > (1-p)(N-2)+1+1 \\ &\iff q(N-2) > (1-p)(N-2)+1 \\ &\iff N\big(p+q-1\big) > \big(p+q-1\big)2+1 \\ &\iff N > 2+\frac{1}{p+q-1} \end{aligned}$$

Proof of Lemma 3.

Consider the map $\Psi^{\hat{m}} : \Omega \to \Omega$, and define an equivalence relation $\sim_\Psi$ on Ω by, $\omega' \sim_\Psi \omega'' \iff \Psi^{\hat{m}}(\omega') = \Psi^{\hat{m}}(\omega'')$. The equivalence relation $\sim_\Psi$ partitions Ω into the quotient set $\Omega/\sim_\Psi$, with $\sim_\Psi$-classes $\omega\big|_{\sim_\Psi}$. Define an order, $\leqslant_{\sim_\Psi}$, on the $\sim_\Psi$-classes in the following way: $\omega'\big|_{\sim_\Psi} \leqslant_{\sim_\Psi} \omega''\big|_{\sim_\Psi} \iff \Psi^{\hat{m}}(\omega') \leqslant_a \Psi^{\hat{m}}(\omega'')$. Now fix $\omega' <_a \omega'' <_a \omega'''$ with $\omega' \sim_\Psi \omega'''$. Then, by defining the *natural mapping* $\jmath_{\sim_\Psi} : \Omega \to \Omega/\sim_\Psi$, it must be that

$$\jmath_{\sim_\Psi}(\omega') \leqslant_{\sim_\Psi} \jmath_{\sim_\Psi}(\omega'') \leqslant_{\sim_\Psi} \jmath_{\sim_\Psi}(\omega''') = \jmath_{\sim_\Psi}(\omega')$$

Proof of Lemma 5.

Cases 1 and 2 are analogous. I prove only the first.
First, note that since $\Omega^{A,b\succeq a}$ is an down-set and that $\Psi(\omega) \leqslant_a \omega$, for all $\omega \in \Omega^{A,b\succeq a}$, then the constraint that $\omega^\star \in (\Omega\backslash\Omega^{A,b\succeq a})_-$ is trivially satisfied since $(\Omega\backslash\Omega^{A,b\succeq a})$ is itself an up=set. Second, recall that $\Omega^{A,b\succ a} \subseteq \Omega^{A,b\succeq a}$, with equality generically. Even non-generically, for sufficiently large m, we have $\Psi^m(\omega) = \omega_{bb}$, for all $\omega \in \Omega^{A,b\succeq a}\backslash(n_a^A-1,0)$.
For any path $h' \in H(\omega, \Omega\backslash\Omega^{A,b\succeq a})$, define

$$g(h') := \#\Big\{\omega' \in \Omega^{A,b\succeq a} \,\Big|\, \exists \omega'' \in \Omega, (\omega' \to \omega'') \in h'\Big\}$$

By the uniqueness of the vertices in a path, $1 \leq g(h') \leq |\Omega^{A,b\succeq a}|$. Define

$$C(\omega;\gamma) := \min\Big\{c_\Psi(h') \,\Big|\, h' \in H(\omega, \Omega\backslash\Omega^{A,b\succeq a}),\ \text{and } g(h') \leq \gamma\Big\}$$

We show that $C(\omega;\gamma)$ is attained by $h^\star$ for all $\gamma \leq |\mathcal{V}(\omega_{bb})|$.
The proof is by induction. First consider the case where $\gamma = 1$.
Clearly ω is the only state in $\Omega^{A,b\succeq a}$ which can be the initial state of an edge in h', so its immediate successor, ω^o lies in $\Omega\backslash\Omega^{A,b\succeq a}$. For each $\omega \in \Omega^{A,b\succeq a}$, define the following set, $\mathcal{D}(\omega) := \Big\{\omega' \,\Big|\, \omega' \in \text{argmin}_{\hat{\omega}\in(\Omega\backslash\Omega^{A,b\succeq a})_-} \|\hat{\omega}, \Psi(\omega)\|\Big\}$. By the definition of $(\Omega\backslash\Omega^{A,b\succeq a})_-$,we have that $\omega^o \geqslant_a \hat{\omega}$, for some $\hat{\omega} \in (\Omega\backslash\Omega^{A,b\succeq a})_-$. So the cost of this path h' is at least $c_\Psi(h') = \|\omega^o, \Psi(\omega)\| \geq \|\omega^\star, \Psi(\omega)\| = c_\Psi(h^\star)$. Clearly then,

$$c_\Psi(h^\star) = C(\omega;1)$$

Now for the inductive step.
Assume that for some γ, $2 \leq \gamma \leq |(\Omega^{A,b\succeq a})| - 1$, and for all $\omega \in \Omega^{A,b\succeq a}$ we have that $c_\Psi(h^\star) = C(\omega;\gamma)$. Fix $\omega \in \Omega^{A,b\succeq a}$. Let ω^o be the immediate successor of ω in some path h^o that is cost minimizing over all paths $h' \in H(\omega, \Omega\backslash\Omega^{A,b\succeq a})$, satisfying $g(h') \leq \gamma + 1$, i.e. that $c(h^o) = C(\omega;\gamma+1)$. It is clear that $\omega^o \in \Omega^{A,b\succeq a} \cup \mathcal{D}(\omega)$, since otherwise

$$\begin{aligned} c_\Psi(h^o) &\geq \|\omega^o, \Psi(\omega)\| \\ &> \underset{\hat{\omega}\in(\Omega\backslash\Omega^{A,b\succeq a})_-}{\text{argmin}} \|\hat{\omega}, \Psi(\omega)\| \\ &= \|\omega^\star, \Psi(\omega)\| \\ &= c_\Psi(h^\star) \end{aligned}$$

If $\omega^o \in \text{argmin}_{\hat{\omega}\in(\Omega\backslash\Omega^{A,b\succeq a})_-} \|\hat{\omega}, \Psi(\omega)\|$, then we have $c_\Psi(h^o) = c_\Psi(h^\star)$. So we will assume $\omega^o \in \Omega^{A,b\succeq a}$.
By the induction hypothesis, we have that the path

$$h'' = \Big\{(\omega^o \to \omega^{\star\star})\Big\}$$

where $\omega^{\star\star} \in \text{argmin}_{\hat{\omega}\in(\Omega\backslash\Omega^{A,b\succeq a})_-} \|\hat{\omega}, \Psi(\omega^o)\|$, is the path of minimum cost in $H(\omega^o, \Omega\backslash\Omega^{A,b\succeq a})$ in γ or fewer steps. i.e.

$$c(h'') = C(\omega^o;\gamma)$$

Take the edge $(\omega \to \omega^o)$ and glue this on to the initial node of h'', ω^o. Call this new path h'''. We have shown that $c(h''') = c(h^o)$ where

$$h''' := \{(\omega \to \omega^o)\} \cup \{\omega^o \to \omega^{\star\star}\}$$

Let us now show that $c_\Psi(t''') \geq c_\Psi(t^\star)$. There are two cases to consider:

1. $\omega \perp_a \omega^o$.
   It must be that either $\omega^o \leqslant_a \omega$ or $\omega <_a \omega^o$.

   - $\omega^o \leqslant_a \omega$.
     We have that $\Psi(\omega^o) \leqslant_a \Psi(\omega)$. Therefore, $\mathcal{D}(\omega) \subseteq \mathcal{D}(\omega^o)$, so we can choose $\omega^{\star\star} = \omega^\star$. Hence $c_\Psi(h''') = \|\Psi(\omega), \omega^o\| + \|\omega^{\star\star}, \Psi(\omega^o)\| \geq \|\omega^{\star\star}, \Psi(\omega^o)\| = \|\omega^\star, \Psi(\omega^o)\| \geq \|\omega^\star, \Psi(\omega)\| = c(h^\star)$.

- $\omega <_a \omega^o$.
  It must be that $\Psi(\omega) \leqslant_a \Psi(\omega^o) <_a \omega^o$. So, while now $\mathcal{D}(\omega^o) \subseteq \mathcal{D}(\omega)$, we can again choose $\omega^{\star\star} = \omega^\star$. In this case

$$\begin{aligned}
c_\Psi(h''') &= \|\Psi(\omega), \omega^o\| + \|\omega^\star, \Psi(\omega^o)\| \\
&= \big(\omega^o_A - [\Psi(\omega)]_A\big) + \big(\omega^o_B - [\Psi(\omega)]_B\big) + \big(\omega^\star_A - [\Psi(\omega^o)]_A\big) + \big(\omega^\star_B - [\Psi(\omega^o)]_A\big) \\
&= \big(\omega^o_A - [\Psi(\omega^o)]_A\big) + \big(\omega^o_B - [\Psi(\omega^o)]_B\big) + \big(\omega^\star_A - [\Psi(\omega)]_A\big) + \big(\omega^\star_B - [\Psi(\omega)]_A\big) \\
&= \|\omega^o, \Psi(\omega^o)\| + \|\omega^\star, \Psi(\omega)\| \\
&\geq \|\omega^\star, \Psi(\omega)\| \\
&= c_\Psi(h^\star)
\end{aligned}$$

2. $\omega \parallel_a \omega^o$.

   There are three cases to consider:

- $\omega^o \leqslant_a \Psi(\omega)$.
  This implies that $\Psi(\omega^o) \leqslant_a \Psi(\omega)$, and so we have that $\mathcal{D}(\omega) \subseteq \mathcal{D}(\omega^o)$. Again, we can choose $\omega^{\star\star} = \omega^\star$, and so $c_\Psi(h''') = \|\Psi(\omega), \omega^o\| + \|\omega^{\star\star}, \Psi(\omega^o)\| \geq \|\omega^{\star\star}, \Psi(\omega^o)\| = \|\omega^\star, \Psi(\omega^o)\| \geq \|\omega^\star, \Psi(\omega)\| = c_\Psi(h^\star)$.
- $\omega^o >_a \Psi(\omega)$.
  It is clear that $\operatorname{argmin}_{\hat\omega \in (\Omega \setminus \Omega^{A, b \succeq a})_-} \|\hat\omega, \omega^o\| \subseteq \mathcal{D}(\omega) \cap \mathcal{D}(\omega^o)$. So, again setting $\omega^{\star\star} = \omega^\star$, we have that

$$\begin{aligned}
c_\Psi(h''') &= \|\Psi(\omega), \omega^o\| + \|\omega^\star, \Psi(\omega^o)\| \\
&= \big(\omega^o_A - [\Psi(\omega)]_A\big) + \big(\omega^o_B - [\Psi(\omega)]_B\big) + \big(\omega^\star_A - [\Psi(\omega^o)]_A\big) + \big(\omega^\star_B - [\Psi(\omega^o)]_A\big) \\
&= \big(\omega^o_A - [\Psi(\omega^o)]_A\big) + \big(\omega^o_B - [\Psi(\omega^o)]_B\big) + \big(\omega^\star_A - [\Psi(\omega)]_A\big) + \big(\omega^\star_B - [\Psi(\omega)]_A\big) \\
&= \|\omega^o, \Psi(\omega^o)\| + \|\omega^\star, \Psi(\omega)\| \\
&\geq \|\omega^\star, \Psi(\omega)\| \\
&= c_\Psi(h^\star)
\end{aligned}$$

- $\omega^o \parallel_a \Psi(\omega)$.
  Without loss of generality we can assume that $[\Psi_\omega]_A < [\omega^o]_A$ and $[\Psi_\omega]_B > [\omega^o]_B$. If $\Psi(\omega^o)$ and $\Psi(\omega)$ cannot be ordered but $\mathcal{D}(\omega) \cap \mathcal{D}(\omega^o) \neq \emptyset$, then the result is immediate. So assume $\mathcal{D}(\omega) \cap \mathcal{D}(\omega^o) \neq \emptyset$, with $[\Psi(\omega)]_A < [\Psi(\omega^o)]_A$ and $[\Psi(\omega)]_B > [\Psi(\omega^o)]_B$. In this case $c_\Psi(\omega, \omega^o) \geq [\omega^o]_A - [\Psi(\omega)]_A > [\Psi(\omega^o)]_A - [\Psi(\omega)]_A > \big[\big(\mathcal{D}(\omega)\big)_{SE}\big]_A - [\Psi(\omega)]_A = \|\omega^\star, \Psi(\omega)\|$, where the first inequality follows by assumption, the second by monotonicity, and the third since $\mathcal{D}(\omega) \cap \mathcal{D}(\omega^o) = \emptyset$.

Proof of Theorem 2.

Partition the set $\Omega^{A, a \succ b} \cap \Omega^{B, b \succ a}$ into $\{\Omega_{ab}, \bar\Omega_{ab}\}$, where

- $\Omega_{ab} := \{\omega \mid [\omega]_A \geq n^A_a\} \cap \{\omega \mid [\omega]_B \leq N^B - n^B_b\})$
- $\bar\Omega_{ab} := \big(\Omega^{A, a \succ b} \cap \Omega^{B, b \succ a}\big) \setminus \Omega_{ab}$.

Lemma 8 below is key in computing the minimum of $c_\Psi(h')$ and $c_\Psi(h'')$ over all $h' \in H(\omega_{ab}, \Omega^{A,b\succeq a})$ and all $h'' \in H(\omega_{ab}, \Omega^{B,a\succeq b})$. Lemma 9 shows that when both groups adopt at constant rates, the partition of Ω into regions of preference is always closely related to the basins of attraction. For reasons of simplicity, Lemma 9 is stated for the generic case only. Both proofs are omitted. That of Lemma 8 proceeds along similar lines to Lemma 5, while that of Lemma 9 is simple.

Lemma 8. *Suppose $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, and let Ψ be a monotonic Group-Darwinian adjustment process such that both groups adapt at constant rates. Then,*

1. *For all $\omega \in \Omega_{ab}$, the minimum of $c_\Psi(h')$ over all paths $h' \in H(\omega, \Omega^{B,a\succeq b})$ is attained by*

$$h^{\star\star} := \begin{cases} \Big\{\big(\omega \to \Psi(\omega)\big), \ldots, \big(\Psi^{\hat{m}-1}(\omega) \to \Psi^{\hat{m}}(\omega)\big)\Big\} \cup \Big\{(\omega_{ab}, \omega^\star)\Big\}, \\ \{(\omega \to \omega^{\star\star})\}, \\ \Big\{\big(\omega \to \Psi(\omega)\big), \ldots, \big(\Psi^{k-2}(\omega) \to \Psi^{k-1}(\omega)\big)\Big\} \cup \Big\{(\Psi^{k-1}(\omega) \to \omega^{\star\star\star})\Big\}, \end{cases}$$

$$\text{if} \begin{cases} [\Psi^m(\omega')]_A + [\Psi^m(\omega')]_B \leq N^A, \text{ for all } m \\ [\Psi(\omega)]_A + [\Psi(\omega)]_B \geq N^A \text{ and } \Psi^B \succeq_d \Psi^A \\ [\Psi^m(\omega')]_A + [\Psi^m(\omega')]_B \geq N^A, \text{ for some } m, \text{ and } \Psi^A \succ_d \Psi^B \end{cases}$$

   where,

   - $\omega^\star \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{B,a\succeq b}} \|\omega_{ab}, \hat{\omega}\|$
   - $\omega^{\star\star} \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{B,a\succeq b}} \|\Psi(\omega), \hat{\omega}\|$
   - $\omega^{\star\star\star} \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{B,a\succeq b}} \|\Psi^k(\omega), \hat{\omega}\|$
   - $k = \min_m \min_{\hat{\omega} \in \Omega^{B,a\succeq b}} \|\Psi^m(\omega), \hat{\omega}\|$

2. *For all $\omega \in \Omega_{ab}$, the minimum of $c_\Psi(h'')$ over all paths $h'' \in H(\omega, \Omega^{A,b\succeq a})$ is attained by*

$$h^{\star\star} := \begin{cases} \Big\{\big(\omega \to \Psi(\omega)\big), \ldots, \big(\Psi^{\hat{m}-1}(\omega) \to \Psi^{\hat{m}}(\omega)\big)\Big\} \cup \Big\{(\omega_{ab} \to \omega^\star)\Big\}, \\ \{(\omega \to \omega^{\star\star})\}, \\ \Big\{\big(\omega \to \Psi(\omega)\big), \ldots, \big(\Psi^{k-2}(\omega) \to \Psi^{k-1}(\omega)\big)\Big\} \cup \Big\{(\Psi^{k-1}(\omega) \to \omega^{\star\star\star})\Big\}, \end{cases}$$

$$\text{if} \begin{cases} [\Psi^m(\omega')]_A + [\Psi^m(\omega')]_B \geq N^A, \text{ for all } m \leq \hat{m} \\ [\Psi(\omega)]_A + [\Psi(\omega)]_B \leq N^A \text{ and } \Psi^A \succeq_d \Psi^B \\ [\Psi^m(\omega')]_A + [\Psi^m(\omega')]_B \leq N^A, \text{ for some } m, \text{ and } \Psi^B \succ_d \Psi^A \end{cases}$$

   where,

   - $\omega^\star \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{A,b\succeq a}} \|\omega_{ab}, \hat{\omega}\|$
   - $\omega^{\star\star} \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{A,b\succeq a}} \|\Psi(\omega), \hat{\omega}\|$
   - $\omega^{\star\star\star} \in \operatorname{argmin}_{\hat{\omega} \in \Omega^{A,b\succeq a}} \|\Psi^k(\omega), \hat{\omega}\|$.
   - $k = \min_m \min_{\hat{\omega} \in \Omega^{A,b\succeq a}} \|\Psi^m(\omega), \hat{\omega}\|$

Lemma 9. *Suppose $E(\mathcal{G}) = \{\omega_{bb}, \omega_{ab}, \omega_{aa}\}$, and let Ψ be a monotonic Group-Darwinian adjustment process such that both groups adapt at constant rates. Then,*

$$\begin{array}{lll} \text{if } \Psi^A \succ_d \Psi^B, & \text{then} & \mathcal{V}(\omega_{aa}) \supseteq \Omega^{B,a\succeq b}, \ \mathcal{V}(\omega_{bb}) = \Omega^{A,b\succeq a} \text{ and } \mathcal{V}(\omega_{ab}) \subseteq \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a} \\ \text{if } \Psi^B \succ_d \Psi^A, & \text{then} & \mathcal{V}(\omega_{aa}) = \Omega^{B,a\succeq b}, \ \mathcal{V}(\omega_{bb}) \supseteq \Omega^{A,b\succeq a} \text{ and } \mathcal{V}(\omega_{ab}) \subseteq \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a} \\ \text{if } \Psi^B \sim_d \Psi^A, & \text{then} & \mathcal{V}(\omega_{aa}) = \Omega^{B,a\succeq b}, \ \mathcal{V}(\omega_{bb}) = \Omega^{A,b\succeq a} \text{ and } \mathcal{V}(\omega_{ab}) = \Omega^{A,a\succ b} \cap \Omega^{B,b\succ a} \end{array}$$

We construct the minimum cost ω-trees for ω_{bb}, ω_{ab}, and ω_{aa}. Denote these trees of minimum cost by $\tau^{\star}_{\omega_{bb}}$, $\tau^{\star}_{\omega_{ab}}$, and $\tau^{\star}_{\omega_{aa}}$ respectively. Since the construction of $\tau^{\star}_{\omega_{bb}}$, parallels exactly the construction of $\tau^{\star}_{\omega_{aa}}$, we construct only the ω-trees $\tau^{\star}_{\omega_{ab}}$ and $\tau^{\star}_{\omega_{aa}}$.

Since $\mathcal{V}(\omega_{bb})$, $\mathcal{V}(\omega_{ab})$, and $\mathcal{V}(\omega_{aa})$ are convex (Lemma 3), we have that for every pair $\omega' \in \mathcal{V}(\omega_{bb})$ and $\omega'' \in \mathcal{V}(\omega_{aa})$ such that $\omega' \perp \omega''$, there exists $\hat{\omega} \in \mathcal{V}(\omega_{ab})$ such that $\omega' <_a \hat{\omega} <_a \omega''$. Furthermore, both groups adapt at equal rates, so by lemma 9 boundaries and total-boundaries of basins of attraction coincide. We have

$$
\begin{aligned}
(\mathcal{V}(\omega_{bb}))_+ = (\mathcal{V}(\omega_{bb}))_{++} &= \begin{cases} \left\{\omega \mid [\omega]_A + [\omega]_B = n_a^A - 1\right\}, & \text{if genA} \\ \left\{\omega \mid [\omega]_A + [\omega]_B = n_a^A - 1\right\} \setminus (n_a^A - 1, 0), & \text{if ngenA} \end{cases} \\
(\mathcal{V}(\omega_{ab}))_- = (\mathcal{V}(\omega_{ab}))_{--} &= \left\{\omega \mid [\omega]_A + [\omega]_B = n_a^A\right\} \\
(\mathcal{V}(\omega_{ab}))_+ = (\mathcal{V}(\omega_{ab}))_{++} &= \left\{\omega \mid [\omega]_A + [\omega]_B = N - n_b^B\right\} \\
(\mathcal{V}(\omega_{aa}))_- = (\mathcal{V}(\omega_{aa}))_{--} &= \begin{cases} \left\{\omega \mid [\omega]_A + [\omega]_B = n_a^B\right\}, & \text{if genB} \\ \left\{\omega \mid [\omega]_A + [\omega]_B = n_a^B\right\} \setminus (n_a^B, 0), & \text{if ngenB} \end{cases}
\end{aligned}
$$

- Construction of $\tau^{\star}_{\omega_{aa}}$.

  For any ω_{aa}-tree, $\tau_{\omega_{aa}}$, define

$$
\begin{aligned}
h'(\omega_{bb}, \omega_{aa}) &= \left\{(\omega' \to \omega'') \in \tau_{\omega_{aa}} \,\middle|\, \omega' = \omega_{bb} \text{ or } \omega' \text{ is a successor of } \omega_{bb}\right\} \\
h'(\omega_{ab}, \omega_{aa}) &= \left\{(\omega' \to \omega'') \in \tau_{\omega_{aa}} \,\middle|\, \omega' = \omega_{ab} \text{ or } \omega' \text{ is a successor of } \omega_{ab}\right\}
\end{aligned}
$$

  Any ω_{aa}-tree, $\tau_{\omega_{aa}}$, must satisfy exactly one of the following four properties,

  1. $\tau^{(1)}_{\omega_{aa}}$: $h'(\omega_{ab}, \omega_{aa}) \subseteq h'(\omega_{bb}, \omega_{aa})$
  2. $\tau^{(2)}_{\omega_{aa}}$: $h'(\omega_{bb}, \omega_{aa}) \subseteq h'(\omega_{ab}, \omega_{aa})$
  3. $\tau^{(3)}_{\omega_{aa}}$: $h'(\omega_{bb}, \omega_{aa}) \cap h'(\omega_{ab}, \omega_{aa}) = \emptyset$
  4. $\tau^{(4)}_{\omega_{aa}}$: $h'(\omega_{bb}, \omega_{aa}) \cap h'(\omega_{ab}, \omega_{aa}) \neq \emptyset$ but neither $\tau^{(1)}_{\omega_{aa}}$ nor $\tau^{(2)}_{\omega_{aa}}$

  We show that $\tau^{\star}_{\omega_{aa}}$ must possess property $\tau^{(1)}_{\omega_{aa}}$.

  By Lemma 8, the minimum of $c_{\Psi}\big(h(\omega_{ab}, \omega_{aa})\big)$ over paths in $H(\omega_{ab}, \omega_{aa})$ is achieved by

$$
h^{\star} = \left\{\big(\omega_{ab} \to (N^A, n_a^B - N^A)\big)\right\} \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \mid \omega' = \Psi^m\big((N^A, n_a^B - N^A)\big) \text{ for some } m \geq 0\Big\}
$$

  where $c_{\Psi}(h^{\star}) = (n_a^B - N^A)$. While $h^{\star}$ above is different for the nongeneric case, it is easy to see that the cost $c_{\Psi}(h^{\star})$ is unchanged.

  Decompose $h'(\omega_{bb}, \omega_{aa})$ into paths h' and h'' where $h' \in H(\omega_{bb}, \Omega \setminus \Omega^{A, b \succeq a})$, and $h'' = h'(\omega_{bb}, \omega_{aa}) \setminus h'$, so that $c_{\Psi}\big(h'(\omega_{bb}, \omega_{aa})\big) = c_{\Psi}(h') + c_{\Psi}(h'')$. By Lemma 5, the minimum of $c_{\Psi}(h')$ is attained by $h^{**} = \{(\omega_{ab} \to \omega^{\star\star})\}$, where $\omega^{\star\star} \in (\mathcal{V}(\omega_{ab}))_-$. Because $\Psi^{\hat{m}}(\omega^{\star\star}) = \omega_{ab}$ for all $\omega^{\star\star} \in \mathcal{V}(\omega_{ab})$, an upper bound for $h'(\omega_{bb}, \omega_{aa})$ is given by

$$
c_{\Psi}(h') + c_{\Psi}(h^{\star}) = n_a^A + (n_b^A - N^A) \tag{21}
$$

  Since $n_a^A \leq N^A$, and $k_A = k_B$, for all $\omega^{\star\star} \in (\mathcal{V}(\omega_{ab}))_-$ and all $m \geq 0$, it must be that $n_a^A \leq [\Psi^m(\omega^{\star\star})]_A + [\Psi^m(\omega^{\star\star})]_B \leq N^A$. If $n_a^A = N^A$, and $\Psi^k(\omega^{\star\star}) \neq \omega_{ab}$ with $\left[\Psi^k(\omega^{\star\star})\right]_A \geq N^A - n_b^B$ for some $k \geq 1$, and nongenericA, then by techniques similar to Lemma 8, the minimum of

$c_\Psi\big(h(\omega_{bb},\omega_{aa})\big)$ over all $H(\omega_{bb},\omega_{aa})$ is attained by

$$\begin{aligned}h''' = &\big\{\big(\omega_{bb}\to(0,n_a^A)\big)\big\}\cup\Big\{\big(\omega'\to\Psi(\omega')\big)\,|\,\omega'=\Psi^m\big((0,n_a^A)\big)\text{ for }0\le m\le k-1\Big\}\\ &\cup\big\{\big(\Psi^{k-1}\big((0,n_a^A)\big)\to\hat\omega\big)\big\}\end{aligned}$$

where $\hat\omega\in\min_{\omega'\in\mathcal{V}(\omega_{aa})}\|\Psi^k\big((0,n_a^A)\big),\omega'\|$. This yields

$$c_\Psi(h''') = n_a^A + (n_b^A - N^A - 1)$$

Otherwise, it follows easily that $\min_m\min_{\hat\omega\in\Omega^{B,a\succeq b}}\|\Psi^m(\omega^{\star\star}),\hat\omega\|$ is attained at $m=\hat m$ and $\hat\omega=(N^A,N^B-n_b^B+1)$, so that the bound in equation 21 binds.

- $\tau^{(1)}_{\omega_{aa}}$: This exactly attains the bound in equation 21.
- $\tau^{(2)}_{\omega_{aa}}$: This exactly attains the bound in equation 21, but only when $n_a^A=N^A$, $\Psi^k(\omega^{\star\star})\neq\omega_{ab}$ with $\big[\Psi^k(\omega^{\star\star})\big]_A\ge N^A-n_b^B$ for some $k\ge1$, and nongenericA all hold. Otherwise there exists $h'=h(\omega_{ab},\omega_{aa})\backslash h(\omega_{bb},\omega_{aa})$ with $c_\Psi(h')>0$.
- $\tau^{(3)}_{\omega_{aa}}$: In this case, it is clear that since $h(\omega_{bb},\omega_{aa})$ can be decomposed into h' and h'' as before, with $h''\cap h^\star\neq\emptyset$, it must be $c_\Psi\big(h(\omega_{bb},\omega_{aa})\big)>c_\Psi(h^{\star\star})$. Coupling this with the fact that the minimum cost path in $H(\omega_{ab},\omega_{aa})$ has cost equal to $c_\Psi(h^{\star\star})=N^B-n_b^B$, yields the desired result.
- $\tau^{(4)}_{\omega_{aa}}$: Follows along similar lines to $\tau^{(3)}_{\omega_{aa}}$.

• Construction of $\tau^\star_{\omega_{ab}}$.
For any ω_{ab}-tree, $\tau_{\omega_{ab}}$, define

$$\begin{aligned}h'(\omega_{bb},\omega_{ab}) &= \big\{(\omega',\omega'')\in\tau_{\omega_{ab}}\,\big|\,\omega'=\omega_{bb}\text{ or }\omega'\text{ is a successor of }\omega_{bb}\big\}\\ h'(\omega_{aa},\omega_{ab}) &= \big\{(\omega',\omega'')\in\tau_{\omega_{ab}}\,\big|\,\omega'=\omega_{aa}\text{ or }\omega'\text{ is a successor of }\omega_{aa}\big\}\end{aligned}$$

By Lemma 5 the path of minimum cost from ω_{bb} to $\Omega\backslash\Omega^{A,b\succeq a}$ is attained by $h'=(\omega_{bb},\omega^\star)$ where $\omega^\star\in(\Omega\backslash\Omega^{A,b\succeq a})_-$. But note in this case that $(\Omega\backslash\Omega^{A,b\succeq a})_-\subseteq\mathcal{V}(\omega_{ab})$. Thus define

$$h^\star=\begin{cases}\{(\omega_{bb}\to\omega^\star)\}\cup\Big\{\big(\omega',\Psi(\omega')\big)\,|\,\omega'=\Psi^m(\omega^\star)\text{ for some }m\ge0\Big\}, & \text{if genA}\\[2ex] \big\{\big(\omega_{bb}\to(n_a^A-1,0)\big)\big\}\cup\big\{\big((n_a^A-1,0)\to(n_a^A,0)\big)\big\}\\ \cup\Big\{\big(\omega',\Psi(\omega')\big)\,|\,\omega'=\Psi^m\big((n_a^A,0)\big)\text{ for some }m\ge0\Big\}, & \text{if ngenA}\end{cases}\tag{22}$$

Both if genericA or ngenericA, $c_\Psi(h^\star)=n_a^A$. A similar analysis shows that $c_\Psi(h'')$ is minimized over all $h''\in H(\omega_{aa},\omega_{ab})$, by $h^{\star\star}$ where

$$h^{\star\star}=\begin{cases}\{(\omega_{aa}\to\omega^{\star\star})\}\cup\Big\{\big(\omega',\Psi(\omega')\big)\,|\,\omega'=\Psi^m(\omega^\star)\text{ for some }m\ge0\Big\}, & \text{if genB}\\[2ex] \big\{\big(\omega_{aa}\to(N^A,N^B-n_b^B+1)\big)\big\}\\ \cup\big\{\big((N^A,N^B-n_b^B+1)\to(N^A,N^B-n_b^B)\big)\big\}\\ \cup\Big\{\big(\omega',\Psi(\omega')\big)\,|\,\omega'=\Psi^m\big((N^A,N^B-n_b^B)\big)\text{ for some }m\ge0\Big\}, & \text{if ngenB}\end{cases}\tag{23}$$

Again, there are 4 mutually exclusive properties that $\tau^\star_{\omega_{ab}}$ may satisfy,

1. $\tau^{(1)}_{\omega_{ab}}$: $h'(\omega_{bb},\omega_{ab}) \cap h'(\omega_{aa},\omega_{ab}) = \emptyset$
2. $\tau^{(2)}_{\omega_{ab}}$: $h'(\omega_{bb},\omega_{ab}) \subseteq h'(\omega_{aa},\omega_{ab})$
3. $\tau^{(3)}_{\omega_{ab}}$: $h'(\omega_{aa},\omega_{ab}) \subseteq h'(\omega_{bb},\omega_{ab})$
4. $\tau^{(4)}_{\omega_{ab}}$: $h'(\omega_{bb},\omega_{ab}) \cap h'(\omega_{aa},\omega_{ab}) \neq \emptyset$ but neither $\tau^{(2)}_{\omega_{aa}}$ nor $\tau^{(3)}_{\omega_{aa}}$

Recall that $\mathcal{V}(\omega_{ab})$ is *sandwiched* between $\mathcal{V}(\omega_{bb})$ and $\mathcal{V}(\omega_{aa})$, in the sense that $\mathcal{V}(\omega_{bb}) \subset \mathcal{V}(\omega_{ab})^{\downarrow}$ and $\mathcal{V}(\omega_{aa}) \subset \mathcal{V}(\omega_{ab})^{\uparrow}$. Thus, unless genA and genB and $\mathcal{V}(\omega_{ab}) = \{\omega \mid [\omega]_A + [\omega]_B = N^A\}$, in which case property $\tau^{(4)}_{\omega_{ab}}$ may hold due to $\mathcal{D}(\omega_{bb}) \cap \mathcal{D}(\omega_{aa}) \neq \emptyset$, then a straightforward geometric argument shows that $\tau^{\star}_{\omega_{ab}}$ must possess property $\tau^{(1)}_{\omega_{ab}}$. Regardless, combining equations 22 and 23 we get

$$c_\Psi(\tau^{\star}_{\omega_{ab}}) = n^A_a + n^B_b$$

Concluding,

$$\begin{aligned}
c_\Psi(\tau^{\star}_{\omega_{bb}}) &= n^B_b + \left(n^A_b - N^B\right) \\
c_\Psi(\tau^{\star}_{\omega_{ab}}) &= n^B_b + n^A_a \\
c_\Psi(\tau^{\star}_{\omega_{aa}}) &= n^A_a + \left(n^B_a - N^A\right)
\end{aligned}$$

Proof of Theorem 3.

Since $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$ when $\Psi^A \sim_d \Psi^B$, it must be that $c_\Psi(\tau^{\star}_{\omega_{ab}}) \leq c_\Psi(\tau^{\star}_{\omega_{ab}})$ and $c_\Psi(\tau^{\star}_{\omega_{ab}}) \leq c_\Psi(\tau^{\star}_{\omega_{ab}})$. From equations (15)-(17), it must be that

$$n^A_a \leq n^A_b - N^B \tag{24}$$
$$n^B_b \leq n^B_a - N^A \tag{25}$$

and so generically

$$n^A_a \leq N^A - n^A_a + 1 \tag{26}$$
$$n^B_b \leq N^B - n^B_b + 1 \tag{27}$$

Similar to the proof of Theorem 2, I construct minimum cost ω-trees of ω_{bb}, ω_{ab}, and ω_{aa} for any constant rate dynamic. Denote these by $\tau^{\star\star}_{\omega_{bb}}$, $\tau^{\star\star}_{\omega_{ab}}$, and $\tau^{\star\star}_{\omega_{aa}}$ respectively. Since the construction of $\tau^{\star\star}_{\omega_{bb}}$, parallels exactly the construction of $\tau^{\star\star}_{\omega_{aa}}$, I construct only $\tau^{\star\star}_{\omega_{ab}}$ and $\tau^{\star\star}_{\omega_{aa}}$.

- Construction of $\tau^{\star\star}_{\omega_{ab}}$.
  The result hinges on the following fact.

  Fact 1. The cost of $\tau^{\star}_{\omega_{ab}}$ from equation (16) cannot be improved upon.

  Proof. $\tau^{\star}_{\omega_{ab}}$ must contain paths $h^{\star}(\omega_{bb}, \omega_{ab})$ and $h^{\star}(\omega_{aa}, \omega_{ab})$.
  Consider $h^{\star}(\omega_{bb}, \omega_{ab})$. By Lemma 6 the first edge must be of the form $(\omega_{bb} \to \omega^{\star})$ where $\omega^{\star} \in \Omega \backslash \Omega^{A, b \succeq a}$. Without loss of generality we can choose $\omega^{\star} = (n^A_a, 0)$, and the remaining edges will be a series of costless transitions to ω_{ab}. Thus, $c_\Psi\big(h^{\star}(\omega_{bb}, \omega_{ab})\big) = n^A_a$. Similarly, the only costly edge of $h^{\star}(\omega_{aa}, \omega_{ab})$ an be chosen to be $(\omega_{aa} \to (N^A, N^B - n^B_b))$, so that $c_\Psi\big(h^{\star}(\omega_{aa}, \omega_{ab})\big) = n^B_b$.
  Now $\tau^{\star\star}_{\omega_{ab}}$ must contain paths $h^{\star\star}(\omega_{bb}, \omega_{ab}) \in H(\omega_{bb}, \omega_{ab})$ and $h^{\star\star}(\omega_{aa}, \omega_{ab}) \in H(\omega_{aa}, \omega_{ab})$. Consider $h^{\star\star}(\omega_{bb}, \omega_{ab})$. It must contain a costly transition out of $\Omega^{A, b \succeq a}$. But the edge $\big(\omega_{bb} \to$

$(n_a^A, 0)\big)$ chosen in $\tau^\star_{\omega_{ab}}$ can be chosen again. So that we can set $h^{\star\star}(\omega_{bb}, \omega_{ab}) = h^\star(\omega_{bb}, \omega_{ab})$. Similarly we can choose, $h^{\star\star}(\omega_{aa}, \omega_{ab}) = h^\star(\omega_{aa}, \omega_{ab})$.
Summarizing, we have that the minimum cost ω_{ab}-tree, $\tau^{\star\star}_{\omega_{ab}}$, has cost

$$c_\Psi(\tau^{\star\star}_{\omega_{ab}}) = n_b^B + n_a^A$$

which is equal to that of $\tau^\star_{\omega_{ab}}$ as in Theorem 2. □

- Construction of $\tau^{\star\star}_{\omega_{bb}}$.
  There are two cases to consider.
  - $\Psi^A \succeq_d \Psi^B$.
    By Lemma 9, when $\Psi^A \succeq_d \Psi^B$, $\mathcal{V}(\omega_{bb}) \subseteq \Omega^{A,b\succeq a}$. Since it is now possible that $(\Omega \backslash \Omega^{B,a\succeq b})_+ \cap \mathcal{V}(\omega_{aa}) \neq \emptyset$, the minimum of $c_\Psi(\hat{h})$ over all $\hat{h} \in H(\omega_{aa}, \mathcal{V}(\omega_{bb}))$ has edges of positive cost $\big\{(\omega_{aa} \to (N^A, N^B - n_b^B))\big\}$ and $\big\{\omega_{ab} \to (n_a^A - 1, 0)\big\}$ which coincide exactly with the minimum cost ω_{bb}-tree from Theorem 2, $\tau^\star_{\omega_{bb}}$, with cost given by equation 15.
  - $\Psi^B \succ_d \Psi^A$.
    Suppose that $\Psi = (\Psi_1^A, \mathcal{B}^B)$, so that Ψ^B dominates Ψ^A maximally. This maximizes $\mathcal{V}(\omega_{bb})$ so that,

$$\mathcal{V}(\omega_{bb}) = \begin{cases} \big(\big\{(n_a^A - 2, N^B)\big\}^\downarrow \cup \big\{(n_a^A - 1, 0)\big\}\big) \cap \Omega^{B,b\succ a}, & \text{if genA} \\ \big(\big\{(n_a^A - 3, N^B)\big\}^\downarrow \cup \big\{(n_a^A - 2, 1)\big\}^\downarrow\big) \cap \Omega^{B,b\succ a}, & \text{if ngenA} \end{cases}$$

    The cost $n_b^B + n_b^A - N^B$, attained in equation 15 is still attainable by the ω_{bb}-tree, $\tau^\star_{\omega_{bb}}$, constructed in Theorem 2, and so $c_\Psi(\tau^\star_{\omega_{bb}})$ is an upper bound for $c_\Psi(\tau^{\star\star}_{\omega_{bb}})$. The other candidate ω_{bb}-tree, $\hat{\tau}_{\omega_{bb}}$, has paths $h(\omega_{ab}, \omega_{bb})$ and $h(\omega_{aa}, \omega_{bb})$, where $h(\omega_{aa}, \omega_{bb}) \subset h(\omega_{aa}, \omega_{bb})$. The only costly transitions along $\hat{\tau}_{\omega_{bb}}$ are given by $(\omega_{ab} \to \hat{\omega})$ and $(\omega_{aa} \to \hat{\hat{\omega}})$, where $\hat{\omega} = \big((\mathcal{V}(\omega_{aa}))_-\big)_{SE}$ and $\hat{\hat{\omega}} = \big((\mathcal{V}(\omega_{bb}))_+\big)_{NW}$. Thus, $\hat{\tau}_{\omega_{bb}}$ has cost given by

$$\begin{aligned} c_\Psi(\hat{\tau}_{\omega_{ab}}) &= c_\Psi(\omega_{ab}, \hat{\omega}) + c_\Psi(\omega_{aa}, \hat{\hat{\omega}}) \\ &= N^B - n_b^B + 1 + \max\big\{N^A - n_a^A + 2, n_b^B\big\} \end{aligned}$$

    There are two cases to consider:
    * $n_b^B \geq N^A - n_a^A + 2$.
      Then,

$$\begin{aligned} c_\Psi(\hat{\tau}_{\omega_{ab}}) &= (N^B - n_b^B + 1) + n_b^B \\ &\geq (N^B - n_b^B + 1) + (N^A - n_a^A + 1) + 1 \\ &\geq n_b^B + n_a^A + 1 \\ &> c_\Psi(\tau^\star_{\omega_{ab}}) \end{aligned}$$

      where the first inequality follows by assumption, the second using (26) and (27).
    * $n_b^B < N^A - n_a^A + 2$.

Then,

$$\begin{aligned} c_\Psi(\hat{\tau}_{\omega_{ab}}) &= (N^B - n_b^B + 1) + (N^A - n_a^A + 1) + 1 \\ &\geq n_b^B + n_a^A + 1 \\ &> c_\Psi(\tau^\star_{\omega_{ab}}) \end{aligned}$$

using (26) and (27).

Proof of Theorem 4.

Consider part 1 when $(N^A, 0) \in \mathcal{V}(\omega_{aa})$.

1. Construction of $\tau^\star_{\omega_{bb}}$.
This is straightforward. The minimum cost path $h \in H(\omega_{bb}, \omega_{aa})$ must involve a transitioning out of $\Omega^{A,b \succeq a}$. Regardless of rates of evolution, it is always the case that $(n_a^A, 0) \in \mathcal{V}(\omega_{aa})$, since clearly there exists $\hat{m}$ such that for all $m \geq \hat{m}$, $\Psi^m\big((n_a^A, 0)\big) \geq (N^A, 0) \in \mathcal{V}(\omega_{aa})$. Futhermore, $(n_a^A, 0) \in \big(\Omega \backslash \Omega^{A,b \succeq a}\big)_-$, and so by Lemma 5, the path of minimum cost from ω_{bb} to $(n_a^A, 0)$ is to transition there immediately, in the generic case, and transition to $(n_a^A - 1, 0)$ and then to $(n_a^A, 0)$ in the non-generic case. Either way, this path $h(\omega_{bb}, (n_a^A, 0))$ has cost of n_a^A. Clearly then, a path of minimum cost from ω_{bb} to ω_{aa} is given by $h^\star := \Big\{(\omega_{bb} \to (n_a^A, 0))\Big\} \cup \Big\{\big(\omega' \to \Psi(\omega')\big) \,\big|\, \omega' = \Psi^m((n_a^A, 0)),\ \text{for some } m \geq 0\Big\}$, with $c_\Psi(h^\star) = n_a^A$.

2. Construction of $\tau^\star_{\omega_{aa}}$.
In the same way that Lemma 5 rested on the observation that $\min_{\hat{\omega} \in \Omega \backslash \Omega^{A,b \succeq a}} \|\Psi^k(\omega), \hat{\omega}\|$ is increasing in k for all $\omega \in \Omega^{A,b \succeq a}$, this subcase of Theorem 4 hinges on the following Lemma. There are two subcases to consider.

Lemma 10. *Suppose $E(\mathcal{G}) = \{\omega_{bb}, \omega_{aa}\}$ and $\omega_{ab} \in \mathcal{V}(\omega_{aa})$. If Ψ is a constant rate dynamic, then for all $\omega \in \mathcal{V}(\omega_{aa}) \cap \Omega^{B,b \succeq a}$, it is the case that*

$$\min_{\hat{\omega} \in \mathcal{V}(\omega_{bb})} \|\omega, \hat{\omega}\| \;\leq\; \min_{\hat{\omega} \in \mathcal{V}(\omega_{bb})} \|\Psi(\omega), \hat{\omega}\| \tag{28}$$

Proof. When $\Psi^A \succeq_d \Psi^B$ the result is immediate, since $\mathcal{V}(\omega_{bb}) = \Omega^{A,b \succeq a}$ by Lemma 9. So with $\Psi^B \succ_d \Psi^A$, there are two cases to consider: $\omega_{ba} \in \mathcal{V}(\omega_{bb})$ and $\omega_{ba} \notin \mathcal{V}(\omega_{bb})$.

- $\omega_{ba} \in \mathcal{V}(\omega_{bb})$.
Clearly $(\mathcal{V}(\omega_{bb}))_+$ can be uniquely decomposed into a collection of column chains $\{c_i\}_{i=1}^n$, where for any $\omega_{i_k} \in c_i$ and $\omega_{j_l} \in c_j \neq c_i$, it must be that $\omega_{i_k} \parallel_a \omega_{j_l}$. Now note that for any $\omega \in \mathcal{V}(\omega_{aa}) \cap \Omega^{B,b \succ a}$, it must be that $\operatorname{argmin}_{\hat{\omega} \in \mathcal{V}(\omega_{bb})} \|\hat{\omega}, \omega\|$ is attained by either $\{\omega_{i_k}\}$ where $\omega_{i_k} \in c_i$ and $[\omega_{i_k}]_B = [\omega]_B$, or by $\{\omega_{i_k}, \omega_{j_l}\}$ where $\omega_{i_k} \in c_i$, $\omega_{j_l} \in c_j \neq c_i$, with $[\omega_{i_k}]_B = [\omega]_B$ and $[\omega_{j_l}]_B = [\omega_{i_k}]_B - 1$ and $[\omega_{j_l}]_A = [\omega_{i_k}]_A + 1$.
For all $\omega \in \mathcal{V}(\omega_{aa}) \cap \Omega^{B,b \succ a}$ with $[\omega]_B - k_B \geq 0$, it must be that $|\,[\omega]_B - [\Psi\omega)]_B\,| > |\,[\omega]_A - [\Psi\omega)]_A\,|$ and hence (28) holds. If $[\omega]_B - k_B \leq 0$, the result is trivial.
We proceed to a contradiction. Suppose for simplicity's sake that $\operatorname{argmin}_{\hat{\omega} \in \mathcal{V}(\omega_{bb})} \|\hat{\omega}, \omega\|$ is attained by $\omega_{i_k} \in c_i$ where $[\omega_{i_k}]_B = [\omega]_B$, and consider the interval $[\omega_{i_k}, \omega]$. Similarly suppose that $\operatorname{argmin}_{\hat{\omega} \in \mathcal{V}(\omega_{bb})} \|\hat{\omega}, \Psi(\omega)\|$ is attained by $\omega_{j_l} \in c_i$ where $[\omega_{j_l}]_B = [\omega]_B$, and consider the interval $[\omega_{j_l}, \Psi(\omega)]$. Now note that if $\|\omega_{j_l}, \Psi(\omega)\| < \|\omega_{i_k}, \omega\|$, then $\Psi^{\leftarrow}(\omega_{j_l}) >_a \omega_{i_k}$, and hence $\omega_{i_k} \notin \mathcal{V}(\omega_{bb})$.

- $\omega_{ba} \in \mathcal{V}(\omega_{aa})$.
  The techniques used for the case where $\omega_{ba} \in \mathcal{V}(\omega_{bb})$ can again be applied to all states in $\left\{\left(N - n_a^B - 1, ((\mathcal{V}(\omega_{bb}))_{++})_{NW}\right)\right\}^{\downarrow} \cap \Omega^{A,a\succ b}$. And again a similar inductive argument to that above can be applied to those states in $\left\{\left(0, ((\mathcal{V}(\omega_{bb}))_{++})_{NW} + 1\right)\right\}^{\uparrow} \cap \Omega^{A,a\succ b}$, and so the result follows naturally.

$\square$

Armed with Lemma 10, the remainder of the proof now follows by a simple counting argument as in Theorem 3.

Proof of Theorem 7.

The proof uses the following two properties of the $\lceil\cdot\rceil$ function. For any $x, y \in \mathbb{R}$,

$$\lceil x\rceil + \lceil -x\rceil = \begin{cases} 0, & \text{if } x \in \mathbb{Z} \\ 1, & \text{if } x \notin \mathbb{Z} \end{cases}$$

$$\lceil x\rceil + \lceil y\rceil - 1 \leq \lceil x + y\rceil$$

It is sufficient to show that $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$, implies ω_{ab} is socially efficient. For $\omega_{ab} \in \Xi(\mathcal{G}, c_\Psi)$, by Theorem 6, it must be that

$$n_b^B + n_a^A = \min\left\{n_b^B + n_b^A - N^B, n_a^A + n_a^B - N^A\right\}$$

or equivalently, that both

$$N^A \leq n_a^B - n_b^B \tag{29}$$
$$N^B \leq n_b^A - n_a^A \tag{30}$$

Now consider equations 29 and 30 for generic parameters of $\mathcal{G}$ (we show only the case where $(1-q)N + (2q-1) \notin \mathbb{Z}$ - the other case follows along similar lines). Using equations 1 - 4, we have

$$\begin{aligned} N^A &\leq n_a^B - n_b^B \\ &= \lceil q(N-2) + 1\rceil - \lceil (1-q)N + (2q-1)\rceil \\ &= \lceil q(N-2) + 1\rceil + \lceil -(1-q)N - (2q-1)\rceil - 1 \\ &\leq \lceil (N-2)(2q-1)\rceil \end{aligned}$$

where the first equality followed by plugging in for n_a^B and n_b^B, the second by the first property of $\lceil\cdot\rceil$ described above, and the final inequality by the second property of $\lceil\cdot\rceil$ along with some algebraic manipulation. Thus inequalities 29 and 30 can be restated as

$$N^A \leq \lceil (N-2)(2q-1)\rceil \tag{31}$$
$$N^B \leq \lceil (N-2)(2p-1)\rceil \tag{32}$$

Now let us check the requirements on ω_{ab} to be socially efficient. Equation 18 can be rearranged as

$$\begin{aligned} N^A &\leq \left\lfloor \frac{1}{q}(N-1)(2q-1) \right\rfloor \\ &\leq \left\lceil \frac{1}{q}(N-1)(2q-1) \right\rceil \end{aligned} \tag{33}$$

Similarly, for ω_{ab} to be socially efficient requires

$$N^B \leq \left\lceil \frac{1}{p}(N-1)(2p-1) \right\rceil \tag{34}$$

Finally, note that inequalities 31 and 32 imply those in 33 and 34.

References

ARTHUR, W. B. (1989): "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *Economic Journal*, 99(394), 116–31.

BERGIN, J., AND B. L. LIPMAN (1996): "Evolution with State-Dependent Mutations," *Econometrica*, 64(4), 943–956.

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

CANALS, J., AND F. VEGA-REDONDO (1998): "Multi-level evolution in population games," *International Journal of Game Theory*, 27(1), 21–35.

ELLISON, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61(5), 1047–1071.

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

ELY, J. (2002): "Local Conventions," *Advances in Theoretical Economics*, 2(1).

FARRELL, J., AND G. SALONER (1985): "Standardization, Compatibility, and Innovation," *RAND Journal of Economics*, 16(1), 70–83.

FOSTER, D., AND P. YOUNG (1990): "Stochastic evolutionary game dynamics," *Theoretical Population biology*, 38, 219–232.

FREIDLIN, M. I., AND A. D. WENTZELL (1998): *Random Perturbations of Dynamical Systems (Grundlehren der mathematischen Wissenschaften)*. Springer Verlag.

KANDORI, M., G. J. MAILATH, AND R. ROB (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61(1), 29–56.

KARLIN, S., AND H. M. TAYLOR (1975): *A First Course in Stochastic Processes, Second Edition*. Academic Press, 2 edn.

KATZ, M. L., AND C. SHAPIRO (1985): "Network Externalities, Competition, and Compatibility," *American Economic Review*, 75(3), 424–40.

KIM, Y.-G., AND J. SOBEL (1995): "An Evolutionary Approach to Pre-Play Communication," *Econometrica*, 63(5), 1181–1193.

LEWIS, D. K. (1969): *Convention: a Philosophical Study*. Harvard University Press.

NEARY, P. R. (2010a): "Competing Conventions in Networked Societies," Discussion paper, University of California San Diego.

——— (2010b): "Multiple Group Games," Discussion paper, University of California San Diego, Department of Economics.

OECHSSLER, J. (1997): "Decentralization and the coordination problem," *Journal of Economic Behavior and Organization*, 32(1), 119 – 135.

——— (1999): "Competition Among Conventions," *Computational & Mathematical Organization Theory*, 5(1), 31–44.

PESKI, M. (2010): "Generalized risk-dominance and asymmetric dynamics," *Journal of Economic Theory*, 145(1), 216–248.

ROBSON, A. J., AND F. VEGA-REDONDO (1996): "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *Journal of Economic Theory*, 70(1), 65 – 92.

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

——— (1996): "The Economics of Convention," *The Journal of Economic Perspectives*, 10(2), 105–122.

——— (2001): *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press.