The Effects of Heterogeneous Constraints on Social Coordination and Network Formation *

Qingchao Zeng †

April 29, 2025

Abstract

In this paper, we present an evolutionary model of coordination and network formation where there are two groups of agents who face either high or low linking constraints on the number of links. We study the agents' choices of actions in the 2×2 coordination game and the set of agents to whom they link. For the static game, we show that both monomorphic states (all agents play the same action) and polymorphic states (agents play different actions) are Nash equilibria. We then study a noisy best response learning dynamics to select among multiple Nash equilibria in the static game. We find that if both low and high constraints are loose, the risk-dominant strategy is selected. In contrast, if both low and high constraints are tight, the payoff-dominant action arises. Moreover, we present that the co-existence of the risk- and payoff-dominant actions can be observed for some game parameters.

JEL Codes: C72; D83; D85

Keywords: Heterogeneity, Coordination, Network Formation, Stochastic Stability.

*This paper is the third chapter of my Ph.D. thesis at the University of Essex.

†Department of Economics, University of Fribourg, Bd. de Pérolles 90, 1700, Fribourg, Switzerland. Email: qingchao.zeng@unifr.ch.

1 Introduction

In various social and economic activities, people often benefit from adopting the same actions or adhering to some common standards (e.g. Latex vs. Microsoft, $C++$ vs. Python, Windows vs. MacOS, etc.). This can be characterised as coordination games, which have two pure Nash equilibria, i.e. payoff-dominant equilibrium and risk-dominant equilibrium (see Harsanyi & Selten 1988). Related literature points out that agents usually coordinate on the same action (see e.g. Kandori et al. 1993; Young 1993; Blume 1993, 1995; or Ellison 1993, 2000, etc.). However, some examples in our real world reveal that it is often the case people do not choose the same action as others, for example, both $C++$ and Python have positive market shares. Thus, it is important to know what drives people to choose different actions and which actions will be selected in the long run.

To solidify the idea, consider a group of students collaborating on a project. A student is better off if she forms a team with somebody using the same software, i.e. either $C++$ or Python. In addition to which software to use, her payoff from this joint project also depends on the choice of her teammates. Therefore, each student has to make two decisions: software and collaborators to maximise their payoffs. This example gives rise to the co-evolutionary model of 2×2 coordination games and network formation (see e.g. Jackson & Watts 2002, Goyal & Vega-Redondo 2005, Staudigl & Weidenholzer 2014). Moreover, although agents have the flexibility to choose whom they interact with, the number of interactions they can maintain is often limited due to constraints on socialising, through e.g. decreasing marginal payoff from socialising or increasing marginal cost of interaction (see e.g. Jackson & Watts 2002, Staudigl & Weidenholzer 2014, Cui & Weidenholzer 2021, or Cui & Shi 2022). Previous work assumes that constraints on interactions are homogeneous for all agents. However, there is empirical evidence in real-life social networks like Twitter revealing that the number of links that agents can support is different (see e.g. Albert et al. 1999 and Kwak et al. 2010). Thus, it seems more realistic to assume that such constraints on interactions are heterogeneous across agents. Therefore, in this paper, we set up a co-evolutionary model of coordination game and network formation, to study how such heterogeneous constraints affect agents' action choices and linking decisions, and further, which action in the coordination game is selected in the long run.[1]

To be more specific, we follow the model of Staudigl & Weidenholzer (2014) but assume that agents face heterogeneous linking constraints. More precisely, we consider a 2×2 coordination game played among a finite number of agents. Every agent makes two choices simultaneously: the action played in the coordination game and the set of agents she plays the game with. Forming links is costly. The payoff of each agent is the sum of payoffs from the coordination game played with each agent she links to, minus the total cost of forming links. We assume that there are two groups of agents who face two different linking constraints: high and low. The size of the *low-constraint group* (i.e. the number of agents in the group) is assumed to be larger than the size of the *high-constraint group*.[2] Linking costs are assumed to be low enough so that in principle any link is beneficial. Thus, it is optimal for each agent to form as many links as her constraint. However, the optimal actions in the coordination game may be different for low- and high-constraint groups.

In fact, our model shows that *polymorphic states* (where agents with different constraints play different actions) may be Nash equilibria for some given game parameters. Specifically, the profiles where agents in the low-constraint group play the payoff-dominant action and agents in the high-constraint group play the risk-dominant action can be Nash equilibria if the low and high constraints are significantly different. To see this point, consider a polymorphic state as described above. Then it may be the case that agents in the low-constraint group focus all of their links on agents playing payoff-dominant action and thus get the highest possible payoff, which is the mechanism that drives the results in Staudigl & Weidenholzer (2014). However, agents in the high-constraint group may lack sufficient potential interaction partners with the payoff-dominant action. Instead, they face a distribution of mixed actions involving both risk-dominant and payoff-dominant actions, such that playing the risk-dominant action may yield a higher expected payoff. This is similar to the mechanism in Goyal & Vega-Redondo (2005) where the complete network is formed and the risk-dominant action does well. Thus, such a polymorphic state can be a Nash equilibrium for some given parameters. However, the other kind of polymorphic state where agents in the low-constraint group play the risk-dominant action and agents in the high-constraint group play the

[1] See also Zeng (2019) for some static properties of Nash equilibria with heterogeneous constraints in 2×2 coordination games and Lu & Shi (2023) for a dynamics analysis of size-dependent minimum effort game.

[2] This assumption has support from the empirical literature (see e.g. Goyal et al. 2006 and Jackson & Rogers 2007) who find that a minority of agents support a large number of links.

payoff-dominant action, can never be a Nash equilibrium. The reason is that if agents with the higher constraint find there are sufficient potential interaction partners with the payoff-dominant action, it has to be also the case for agents with the lower constraint. Thus, it is always profitable for agents in the low-constraint group to deviate from the risk-dominant action. In addition, in line with Goyal & Vega-Redondo (2005) and Staudigl & Weidenholzer (2014), *monomorphic states* (where all agents play the same action) are always Nash equilibria since when all other agents play the same action, an agent will always be better off if she chooses the same action as others.

Further, given the multiplicity of Nash equilibria, we study the co-evolution of the above static game in discrete time to predict which kind of profiles will be selected in the long run. We assume that at each period, agents may receive opportunities to revise their strategies based on a noisy myopic best-response rule. That is, agents choose actions and links that optimise their payoffs against the distribution of actions in the previous period. There is however a probability that agents make mistakes and choose a random strategy. We follow the standard methodology developed by Kandori et al. (1993), Young (1993) and Freidlin & Wentzell (1998) to identify the stochastically stable states as the long-run prediction, which are the states in the support of a unique invariant distribution when the probability of making mistakes approaches zero. Naturally, states that are more robust to mistakes are stochastically stable.

In the first step, we identify the absorbing sets, which are the sets of states that once reached can never be left without mistakes. The literature considering homogeneous constraints (see e.g. Staudigl & Weidenholzer 2014 and Cui & Shi 2022) shows that the absorbing sets consist of only monomorphic states. In contrast, in our model, polymorphic states can also be absorbing if the two constraints on links are significantly different. Following this, we characterise the set of stochastically stable states by comparing the robustness of absorbing states to mistakes. In cases where the low and high constraints are close, the set of stochastically stable states contains only monomorphic states, which is in line with the model in Goyal & Vega-Redondo (2005) and Staudigl & Weidenholzer (2014). More precisely, the payoff-dominant action emerges in the long run if both constraints are small. In contrast, the risk-dominant action will be selected if both constraints are high. Surprisingly, we also find that if the low and the high constraints are significantly different from one another, the polymorphic states where agents in the low-constraint group play the payoff-dominant action and agents in the high-constraint group play the risk-dominant action can also be stochastically stable.

The structure of this paper is as follows. In section 2, we review the related literature. Section 3 outlines our model. In section 4, we characterise the Nash equilibria of the static game. Section 5 presents our results on the set of stochastically stable states for different levels of linking constraints. Section 6 concludes. Formal proofs of our results are relegated to the Appendix.

2 Literature review

This paper adds to the literature on the co-evolution of coordination and network formation games (see e.g. Jackson & Watts 2002, Goyal & Vega-Redondo 2005, Staudigl & Weidenholzer 2014.). Jackson & Watts (2002) considers a model where the network is bilaterally formed based on the concept of pairwise stability provided by Jackson & Wolinsky (1996) and points out that whether risk-dominant or payoff-dominant conventions are stochastically stable depends on the relationships between payoffs in the coordination games and linking costs. Goyal & Vega-Redondo (2005) consider the case where agents non-cooperatively form unilateral links and find that which convention will emerge also depends on the relative level of linking costs to payoffs. As the adjustment process in Goyal & Vega-Redondo (2005) is different to the one used in Jackson & Watts (2002), the precise nature of the relationship between payoffs and linking costs differs too.[3] Goyal & Vega-Redondo (2005) show that agents will coordinate on the risk-dominant action if the linking cost is low, and they will select the payoff-dominant action if the linking cost is high. Further, Staudigl & Weidenholzer (2014) extend Goyal & Vega-Redondo's model by considering homogeneous constraints on the number of links agents can support. Their study shows that if the constraint is low compared to the population, the payoff-dominant action is selected, while the risk-dominant action will be selected in the long run if the constraint is high. Cui & Weidenholzer (2021) consider the effect of lock-in on the selection of conventions based on Staudigl & Weidenholzer (2014)'s model, where agents receive payoffs not only from links they form but also from the links they receive. They show that agents using different actions sometimes can also be stochastically stable. Our model differs from these studies in that agents face heterogeneous constraints on the number of

[3]As argued by Goyal & Vega-Redondo (2005), the main source of this discrepancy lies in the fact that in Goyal & Vega-Redondo (2005) actions and links are chosen simultaneously but follow independent process in Jackson & Watts (2002).

links. Agents are distinguished by two different levels of linking constraints: high and low. We find that such heterogeneous linking constraints will lead to the co-existence of both risk-dominant and payoff-dominant actions.

The most closely related literature to the present work is the paper by Lu & Shi (2023). They also study a co-evolutionary model with heterogeneous constraints on links, featuring a minimum-effort game. They find that all agents will choose high effort levels if everybody faces low constraints, while low effort levels will be chosen if constraints are high for everybody. The coexistence of different effort levels happens if constraints are significantly different. We remark that at first sight the mechanism and results are similar. However, there are some important differences between their model and ours. While they focus on the size-dependent minimum-effort games, we study 2×2 coordination games. In minimum-effort games, agents always want to match the effort level of the lowest of their interaction partners. This implies that agents choosing the high effort level can never interact with agents choosing lower effort levels. In contrast, agents with payoff-dominant actions may interact with those with risk-dominant actions in equilibria of our model. Furthermore, our results emphasise that the coexistence is not driven by the particular best response structures of minimum-effort games, where it is always best to keep the effort level in line with the weakest link, but carry over to games where the best response to a mixed profile depends on the exact distribution of actions. In addition, Bilancini & Boncinelli (2018) also study heterogeneous agents and build a model with two different types of agents, where interactions between different types result in additional costs. They show that when the costs of interactions with different types are high, one type will play the risk-dominant action and the other type will play the payoff-dominant action.

In addition to the literature on coordination and network formation games, there is also a strand of literature where agents can determine their interaction partners by moving among a set of locations or islands (see e.g. Oechssler 1997, Dieckmann 1999, Anwar et al. 2002, Bhaskar & Vega-Redondo 2004, and Pin et al. 2017). when there are restrictions on the mobility between locations, or constraints on the capacity of each location, the co-existence of conventions might be observed. However, the co-existence of conventions depends on the limited interaction between locations. In contrast, in our model agents have the flexibility to interact with anybody.

3 Model

We consider a 2×2 coordination game played among the population of n agents, denoted by $N = \{1, 2, \cdots, n\}$ ($n \geq 3$). Each agent i can choose an action a_i from the action set $\mathcal{A} = \{A, B\}$. The payoff of agent i is given by $u(a_i, a_j)$ when she plays this coordination game against agent j. An agent who chooses action A in the coordination game is called an A-agent. Similarly, B-agents are those who play action B. The payoff matrix of this coordination game is given in the following table.

	A	B
A	(a, a)	(c, d)
B	(d, c)	(b, b)

We assume that $b > c$ and $a > d$, so that strategy (A, A) and (B, B) are two pure strategy Nash-Equilibria. Further, we assume $b > a$ so that (B, B) is the payoff-dominant equilibrium that yields the highest payoff. Moreover, we assume that $a + c > b + d$ so that (A, A) is risk-dominant equilibrium according to Harsanyi & Selten (1988), meaning that A is the best response against an agent who plays both actions with equal probability $(\frac{1}{2}, \frac{1}{2})$. Given all those assumptions, we can simply have $c > d$. Further, we assume $a > c$ such that A-agents prefer playing against A-agents over playing against B-agents. Combining all assumptions together we have the following order of payoffs $b > a > c > d$.

Apart from their actions in the coordination game, agents may also determine the set of agents that they link to. We denote by g_{ij} the link decision to agent j made by agent i, where $g_{ij} = 1$ denotes that agent i forms a link to agent j and otherwise $g_{ij} = 0$. We consider the case where links are unilaterally formed, i.e. agent i decides on the link g_{ij} and agent j does not have a say in this link.[4] We assume that agents cannot link to themselves, i.e. $g_{ii} = 0$. Agent i's linking strategy g_i can be defined as a n-dimensional vector, i.e. $g_i = (g_{i1}, g_{i2}, \cdots, g_{in}) \in \mathcal{G}_i = \{0, 1\}^n$. The out-degree of agent i is denoted by $d_i^{out} = \sum_j g_{ij}$, i.e. the number of links that agent i forms. The network formed by all agents is denoted by $g = (g_i)_{i \in N}$. Agent i's pure strategy s_i includes her action choice $a_i \in \mathcal{A}$ and linking strategy $g_i \in \mathcal{G}_i$, i.e. $s_i = (a_i, g_i) \in \mathcal{A} \times \mathcal{G}_i = \mathcal{S}_i$. Further, a

[4] There is also some literature considering bilateral links (e.g. Jackson & Wolinsky 1996, Dutta & Mutuswami 1997 or Jackson & Watts 2002) where forming a link requires the consents of both parties.

strategy profile is denoted by $s = (s_1, s_2, \cdots, s_n) \in \Pi_{i \in N} \mathcal{S}_i = \mathcal{S}$.

Agents play the coordination game only with those agents they link to. We assume that the payoff generated by the coordination game only goes to the agent who forms the link. The agent who is passively linked gets zero from the coordination game played.[5] Forming links is costly and the cost is denoted by γ. The total payoff of an agent is given by the sum of payoffs she receives by playing the coordination game with each agent she links to, minus the total cost incurred by forming those links. Thus, given a strategy profile $s = (s_i)_{i \in N}$, the total payoff for agent i is given by

$$U_i(s_i, s_{-i}) = \sum_{j=1}^{n} g_{ij} \cdot u_i(a_i, a_j) - \gamma \cdot d_i^{out}. \tag{1}$$

We focus on a case where the number of links that agent i can support is limited by k_i, i.e. $d_i^{out} \leq k_i$ as in Staudigl & Weidenholzer (2014).[6] In addition, we are interested in a scenario where linking constraints are heterogeneous among agents. Particularly, we consider a case where there are two types of agents, one with a lower constraint k^ℓ and the other with a higher constraint k^h, i.e. $k^\ell < k^h$. We define the set of agents with the lower constraint as *low-constraint group*, denoted by N_ℓ with $n_\ell = |N_\ell|$ the number of agents. Similarly, *high-constraint group* is the set of agents with the higher constraint, denoted by N_h with $n_h = |N_h|$. We focus on the case $n_\ell > n_h$ where the low-constraint group is larger than the high-constraint group.

Consider a scenario where the linking cost is low, i.e. $\gamma < d$, so that in principle an agent wants to form links to any other agents regardless of their actions. In this case, agents will form the maximum number of links they can support, i.e. k^ℓ or k^h. Thus, the total payoff function above is equivalent to

$$U_i(s_i, s_{-i}) = \sum_{j=1}^{n} g_{ij} \cdot u_i(a_i, a_j) - \gamma \cdot k_i \tag{2}$$

where $k_i \in \{k^\ell, k^h\}$ is the linking constraint for agent i.

[5]Goyal & Vega-Redondo (2005) and Cui & Weidenholzer (2021) also consider a model where agents receive benefits from passive links as well.

[6]Alternatively. we can think of this assumption as links becoming prohibitively expensive as a certain threshold is crossed.

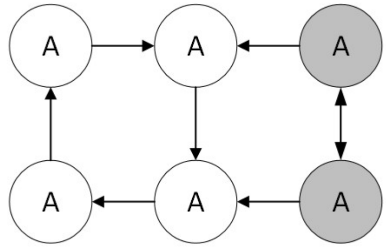


Figure 1: A monomorphic state in $\overrightarrow{AA}$ when $n_\ell = 4, n_h = 2, k^\ell = 1$ and $k^h = 2$.

4 Nash Equilibrium

In our characterization of Nash equilibrium, two types of states play an important role. Firstly, we denote by $\overrightarrow{XX}$ the set of monomorphic states, where $X \in \{A, B\}$. In a monomorphic state, every agent chooses the same action and forms the maximum number of links. More formally, the set of monomorphic states is given by

$$\overrightarrow{XX} = \{s \in S | (a_i = a_j = X) \wedge (d_i^{out} = k^\ell, d_j^{out} = k^h), \forall i \in N_\ell, j \in N_h\}.$$

For example, in a monomorphic state $s \in \overrightarrow{AA}$, all agents play action A, agents in the low-constraint group support k^ℓ links, and agents in the high-constraint group support k^h links. Fig 1 depicts an example of a monomorphic state in $\overrightarrow{AA}$ when $n_\ell = 4, n_h = 2, k^\ell = 1$ and $k^h = 2$.

Secondly, we denote by $\overrightarrow{XY}$ the set of polymorphic states where $X, Y \in \{A, B\}$ and $X \neq Y$. Less formally, in a polymorphic state, all agents in the low-constraint group play one action and all agents in the high-constraint group play the other action. And in terms of linking strategy, all agents exhaust their constraints. Furthermore, agents will first form links to other agents within the same group, and then link to agents in the other group to fill up their remaining slots if any. More formally, $\overrightarrow{XY}$ is defined by

$$\overrightarrow{XY} = \{s \in S | (a_i = X, a_j = Y, a_j \neq a_i) \wedge (d_i^{out} = k^\ell, d_j^{out} = k^h) \wedge (\sum_{i' \in N_\ell} g_{ii'}$$
$$= \min\{k^\ell, n_\ell - 1\}, \sum_{j' \in N_h} g_{jj'} = \min\{k^h, n_h - 1\}), \forall i \in n_\ell, j \in N_h\}.$$

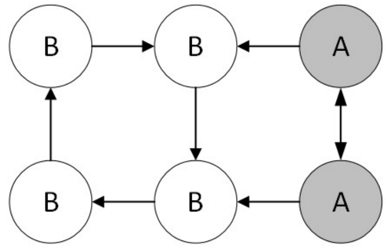


Figure 2: A polymorphic state in $\overrightarrow{BA}$ when $n_\ell = 4, n_h = 2, k^\ell = 1$ and $k^h = 2$.

For example, in a polymorphic state $s \in \overrightarrow{AB}$, all agents in the low-constraint group play the same action A and support k^ℓ links, whereas all agents in the high-constraint group N_h play action B and support k^h links. A- agents link to other A-agents first and then link to B-agents if they still have remaining slots, e.g. if $k^\ell > n_\ell - 1$. Similarly, B- agents link to other B-agents first and then link to A-agents if $k^h > n_h - 1$. Fig 2 depicts an example of a polymorphic state in $\overrightarrow{BA}$ when $n_\ell = 4, n_h = 2, k^\ell = 1$ and $k^h = 2$.

Following the same mechanism as in Staudigl & Weidenholzer (2014), finding the best response can be divided into two steps: First, for each of the two actions, determine the payoff optimizing linking strategy and calculate the payoffs associated with it. This is summarized by the link-optimized payoff functions (for short, the LOPs). And second, compare the LOPs across actions and choose the one with the highest payoff.

We denote by m the number of A-agents at a given strategy profile s. The number of B-agents is thus $n - m$. The LOPs are thus given by

$$v_i(a_i, m) = \max_{g_i \in \mathcal{G}_i} U_i((a_i, g_i), m), \quad \forall i \in N.$$

where $U_i((a_i, g_i), m)$ is agent i's payoff given her strategy $s_i = (a_i, g_i)$ and the number of A-agents m. Consider an agent i whose linking constraint is $k_i \in \{k^\ell, k^h\}$. Given the distribution of actions $(m, n - m)$, her LOP of choosing action A is given by

$$v_i(A, m) = a \cdot \min\{k_i, m - 1\} + c \cdot (k_i - \min\{k_i, m - 1\}) - M \cdot k_i.$$

Intuitively, given the order of payoffs $a > c$, A-agents prefer playing against other A-agents over playing against B-agents. Thus, agent i will first link to other A-agents. Considering different levels between the constraint k_i and the number of other A-agents $m-1$, the maximum number of links to A agents that i could form is $\min\{k_i, m-1\}$. After forming links to A-agents, agent i will then fill her remaining slots $k_i - \min\{k_i, m-1\}$ by linking to B-agents if there are any remaining slots left.

Similarly, the order of payoffs $b > d$ implies that B-agents prefer forming links with other B-agents first. Then they will fill up their remaining slots by linking to A-agents. Note that a B-agents faces $n-(m-1)$ other B-agents if agent i chooses to play action B. Agent i's LOP of choosing action B is thus given by

$$v_i(B, m) = b \cdot \min\{k_i, n-m-1\} + d \cdot (k_i - \min\{k_i, n-m-1\}) - M \cdot k_i.$$

Given the LOPs, we now define the concept of Nash equilibrium in our game. Consider a strategy profile $s \in \mathcal{S}$ and the corresponding distribution of actions $(m, n-m)$. Strategy profile s is a Nash equilibrium if the following two conditions hold:

i) $v_i(A, m) \geq v_i(B, m-1)$ for all i with $a_i = A$;

ii) $v_j(B, m) \geq v_j(A, m+1)$ for all j with $a_j = B$.

We denote by $\mathcal{S}^\star$ the set of Nash equilibria. Given the previous observations, we are now able to state the following proposition which characterizes the set of Nash equilibria. In particular, Nash equilibria correspond to the monomorphic states and the polymorphic states.

Proposition 4.1. *There exist two thresholds* $\overline{k^\ell} = \frac{(b-d)(n_\ell-1)-(a-c)n_h}{(c-d)}$ *and* $\underline{k^h} = \frac{(b-d)n_\ell-(a-c)(n_h-1)}{(c-d)}$, *such that:*

i) if $k^\ell \leq \overline{k^\ell}$ *and* $k^h \geq \underline{k^h}$, *then* $\mathcal{S}^\star = \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{BA}$*;*

ii) if $k^\ell > \overline{k^\ell}$ *or* $k^h < \underline{k^h}$, *then* $\mathcal{S}^\star = \overrightarrow{AA} \cup \overrightarrow{BB}$.

The proof to Proposition 4.1 proceeds using a series of lemmas. Note that each agent's strategy consists of two parts: action choice and linking choice. First, we prove that only monomorphic states and polymorphic states can potentially be a Nash equilibrium, i.e. $\mathcal{S}^\star \subseteq \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{AB} \cup \overrightarrow{BA}$. Next, we show that any strategy profile in $\overrightarrow{AB}$ is not a Nash equilibrium. That is, a Nash equilibrium cannot be a state where agents in the low-constraint group play the risk-dominant action and agents in the high-constraint group play the payoff-dominant action. Then, we prove that for any k^ℓ and k^h, monomorphic states are always Nash equilibria. In the last step, we prove that a strategy profile in $\overrightarrow{BA}$ is Nash equilibrium if and only if $k^\ell \leq \overline{k^\ell}$ and $k^h \geq \underline{k^h}$.

Lemma 1. *If $s \notin \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{AB} \cup \overrightarrow{BA}$, then s is not a Nash equilibrium.*

Intuitively, since agents in the same group have the same constraints, they face the same situation. This implies that whenever it is optimal for one agent to stay at her action, then it is also optimal for agents with the other action to switch. It follows that all agents in the same group have to choose the same action in a Nash equilibrium.

The next lemma establishes that all monomorphic states are Nash equilibria.

Lemma 2. *$\overrightarrow{AA} \cup \overrightarrow{BB} \subset \mathcal{S}^\star$ for any k^ℓ, k^h and n.*

The proof of Lemma 2 is straightforward. First, consider a strategy profile s in the monomorphic set $\overrightarrow{AA}$. The corresponding distribution of actions is $(n, 0)$. Since there are only A-agents, none of them will deviate from playing action A as switching to B will lower their payoff per link by $a - d$. Additionally, no agent has incentives to form fewer links since each link yields $a - \gamma$, which is strictly positive. Thus, no one wants to deviate either from her current action choice or from her linking choice. Therefore, s is a Nash equilibrium. Following the same argument as $s \in \overrightarrow{AA}$, we can also prove that any strategy profile $s \in \overrightarrow{BB}$ is also a Nash equilibrium.

Note that coordinating on the same action always yields a higher payoff than not coordinating. when all agents choose the same action, no one has incentives to switch to the other action. The next two lemmas extend the discussion on Nash equilibrium to the polymorphic states.

Lemma 3. *No state $s \in \overrightarrow{AB}$ is a Nash equilibrium.*

Intuitively, independent of the sizes of the two groups, a Nash equilibrium cannot be a state where agents in the low-constraint group choose the risk-dominant action and agents in the high-constraint group choose the payoff-dominant action. If agents in the high-constraint group choose the payoff-dominant action, it implies that there are sufficient B-agents around for agents with the higher constraint. Thus, for agents with the lower constraint, the number of B-agents is also sufficient. Their best response therefore is choosing the payoff-dominant action B.

The following lemma establishes that a polymorphic state $s \in \overrightarrow{BA}$ could be a Nash equilibrium for some constraints k^ℓ and k^h.

Lemma 4. *$\overrightarrow{BA} \subset \mathcal{S}^\star$ iff $k^\ell \leq \overline{k^\ell}$ and $k^h \geq \underline{k^h}$.*

Lemma 4 provides us with conditions for the co-existence of the risk-dominant action and the payoff-dominant action in a Nash equilibrium. Such a Nash equilibrium is characterized by agents in the low-constraint group choosing the payoff-dominant action, and agents in the high-constraint group choosing the risk-dominant action. We provide two examples to develop intuition for our findings.

Example 3.1. Figure 2 depicts a polymorphic state in $\overrightarrow{BA}$ where $n_\ell = 4, n_h = 2, k^\ell = 1$ and $k^h = 2$. For any payoffs (a, b, c, d) fulfilling our assumptions, one can check that $k^\ell \leq \frac{(b-d)\cdot 3-(a-c)\cdot 2}{c-d}$ and $k^h < \frac{(b-d)\cdot 4-(a-c)}{c-d}$ hold, so that the second condition for polymorphic equilibrium is violated.[7] Therefore, there exists no polymorphic equilibrium. To develop intuition, consider the strategy profile depicted in Figure 2. If agents in the high-constraint group choose action B, their optimal linking choice is to link to two B-agents and they will get $2 \cdot b$ by doing so. If they choose action A, the highest payoff is $a + c$ by linking to one A-agent and one B-agent. Since $b > a > c > d$, we have $2 \cdot b > a + c$, then agents in the high-constraint group will always switch to action B.

Example 3.1 highlights that if the conditions identified in Lemma 4 fail, then there cannot be polymorphic equilibrium. Intuitively, while the low-constraint group has a small enough constraint so that choosing B is optimal, the constraint of agents in the high-constraint group is so small that choosing B would be optimal for them too. If they had a larger constraint, it may be optimal for them to stay with A as the following example shows.

Example 3.2. Figure 3 depicts another example of a polymorphic state in $\overrightarrow{BA}$. Assume that $a+3c \geq 3b+d$.[8] Agents in the low-constraint group will stay with B provided $b > a$. Furthermore, agents in the high-constraint group will stay with A provided $a + 3c \geq 3b + d$. To see this, note that each agent in N_h forms three links with B-agents and one link with the other A-agent. By playing action A, she gets $3c$ from playing against all B-agents and a from playing against the

[7] Note that $b > a > c > d$. Inequality $k^\ell \leq \frac{(b-d)\cdot 3-(a-c)\cdot 2}{c-d}$ holds since that $\frac{(b-d)\cdot 3-(a-c)\cdot 2}{c-d} = \frac{(b-d)+(b-a+c-d)\cdot 2}{c-d} > \frac{(b-d)}{c-d} > 1 = k^\ell$. Inequality $k^h < \frac{(b-d)\cdot 4-(a-c)}{c-d}$ holds since that $\frac{(b-d)\cdot 4-(a-c)}{c-d} = \frac{(b-d)\cdot 3+(b-a+c-d)}{c-d} > \frac{(b-d)\cdot 3}{c-d} > 3 > 2 = k^h$.

[8] One can check that $a + 3c \geq 3b + d$ is plausible, e.g. $(a, b, c, d) = (10, 11, 9, 1)$.

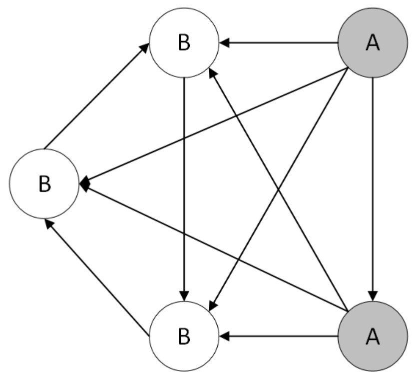


Figure 3: A polymorphic state in $\overrightarrow{BA}$ for $n_\ell = 3, n_h = 2, k^\ell = 1$ and $k^h = 4$, which is a Nash equilibrium in a coordination game if $a + 3c \geq 3b + d$. White circles represent four agents in the low-constraint group. Grey circles represent two agents in the high-constraint group.

other A-agent. If she switches to B, she gets $3b$ from playing against all B-agents and d from playing against the other A-agent. Given that $a + 3c \geq 3b + d$, she will stay with action A. Thus, the state depicted in Figure 3 is a Nash equilibrium.[9]

5 Myopic best response learning

As we have seen in the previous section, our model may feature a multiplicity of equilibria. To assess which of these equilibria is most likely to arise in the long run, we consider a model of myopic best response learning where agents make occasional mistakes in the spirit of Kandori et al. (1993), Young (1993) and Ellison (1993, 2000). The unperturbed model is defined as follows. In discrete-time $t = 0, 1, 2, 3, \cdots$, each agent may receive the opportunity to revise her strategy (both action and links) with a positive probability $\lambda \in (0, 1)$. This probability is independent among all agents and periods. When getting the opportunity to revise in period t, each agent chooses a strategy that maximizes her payoff in the last period $t-1$. More formally, agent i chooses a strategy in the period t as follows:

$$s_i^t \in arg \max_{s_i \in \mathcal{S}_i} U_i(s_i, s_{-i}^{t-1})$$

[9]If $a + 3c \geq 3b + d$, then $k^\ell \leq \frac{(b-d)\cdot 2-(a-c)\cdot 2}{c-d}$ and $k^h \geq \frac{(b-d)\cdot 3-(a-c)}{c-d}$ hold so that the constraints fulfill the conditions identified in Lemma 4.

where s_{-i}^{t-1} is the strategy profile played by agents except i in the last period $t-1$. If there are multiple best responses, agents choose one of them at random.

In light of our discussion in the previous section, this revision protocol can be analyzed in two steps: i) for each action, agents first determine the optimal linking strategy, and ii) given the optimal linking strategies for both actions, agents then determine which of two actions is optimal. This approach is captured by the LOPs. Formally, an agent i chooses her action in the following way:

i) when $a_i^{t-1} = A$, switch to B if $v_i(B, m^{t-1} - 1) > v_i(A, m^{t-1})$, remain with A if $v_i(B, m^{t-1}-1) < v_i(A, m^{t-1})$, randomize between A and B if $v_i(B, m^{t-1}-1) = v_i(A, m^{t-1})$;

ii) when $a_i^{t-1} = B$, switch to A if $v_i(A, m^{t-1} + 1) > v_i(B, m^{t-1})$, remain with B if $v_i(A, m^{t-1}+1) < v_i(B, m^{t-1})$, randomize between A and B if $v_i(A, m^{t-1}+1) = v_i(B, m^{t-1})$,

where a_i^{t-1} denotes i's action and m^{t-1} is the number of A-agents in the last period $t-1$.

The revision rule outlined above gives rise to a Markov chain on the state space $\mathcal{S} \equiv \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$. In this context, a state s in the space $\mathcal{S}$ is equivalent to a strategy profile $s = (s_i)_{i \in N}$.

We are interested in sets of states this process converges to. These sets are known as absorbing sets (see e.g. Kandori et al. (1993), Young (1993), Freidlin & Wentzell (1998), and Ellison (2000)). An absorbing set, denoted by $S^{\star\star}$, is a minimum subset of $\mathcal{S}$ such that:

i) for any pair of states $s, s' \in S^{\star\star}$, the probability of a transition from s to s' is positive;

ii) for any two states $s \in S^{\star\star}$ and $s'' \notin S^{\star\star}$, the probability of a transition from s to s'' is zero.

We denote the set of all absorbing sets by $\mathcal{S}^{\star\star}$.

We now proceed to characterize those absorbing sets. By considering various ranges of the game parameters m, n, k^ℓ, and k^h, we have computed the switching thresholds for agents in both groups, that is, we provide conditions on the distribution of actions when agents find it optimal to switch actions and when they will remain at their current actions. These results are summarized in the Table 1. This allows us to have a full characterization of absorbing sets which is presented as the following proposition. [10]

[10]The existence of different classes of absorbing sets in this setting has already been characterized by Zeng (2019). This proposition goes beyond that result by identifying relevant thresholds.

Table 1: Where ”a.s” means that an agent always switches to the other action and ”n.s” means that an agent never switches to the other action.

Switching Thresholds for A-agents			
$v(B,m-1) \geq v(A,m)$	$k^h > k^\ell \geq m-1$ $k^h > k^\ell \geq n-m$	$k^h \geq m-1 > k^\ell$ $k^h > k^\ell \geq n-m$	$m-1 > k^h > k^\ell$ $k^h > k^\ell \geq n-m$
$i \in N_\ell$	$m \leq \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d} + 1 := M_\ell^1$	$m \leq n - \frac{a-d}{b-d}k^\ell := M_\ell^2$	$m \leq n - \frac{a-d}{b-d}k^\ell$
$j \in N_h$	$m \leq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d} + 1 := M_h^1$	$m \leq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d} + 1$	$m \leq n - \frac{a-d}{b-d}k^h := M_h^2$
	$k^h > k^\ell \geq m-1$ $k^h \geq n-m > k^\ell$	$k^h \geq m-1 > k^\ell$ $k^h \geq n-m > k^\ell$	$m-1 > k^h > k^\ell$ $k^h \geq n-m > k^\ell$
$i \in N_\ell$	$a.s.$	$a.s.$	$a.s.$
$j \in N_h$	$m \leq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d} + 1$	$m \leq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d} + 1$	$m \leq n - \frac{a-d}{b-d}k^h$
		$n-m > k^h > k^\ell$	
$i \in N_\ell$	$a.s.$	$a.s.$	$a.s.$
$j \in N_h$	$a.s.$	$a.s.$	$a.s.$
Switching Thresholds for B-agents			
$v(A,m+1) \geq v(B,m)$	$k^h > k^\ell > m$ $k^h > k^\ell > n-m-1$	$k^h > m \geq k^\ell$ $k^h > k^\ell > n-m-1$	$m \geq k^h > k^\ell$ $k^h > k^\ell > n-m-1$
$i \in N_\ell$	$m \geq \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d}$	$m \geq n-1-\frac{a-d}{b-d}k^\ell$	$m \geq n-1-\frac{a-d}{b-d}k^\ell$
$j \in N_h$	$m \geq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d}$	$m \geq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d}$	$m \geq n-1-\frac{a-d}{b-d}k^h$
	$k^h > k^\ell > m$ $k^h > n-m-1 \geq k^\ell$	$k^h > m \geq k^\ell$ $k^h > n-m-1 \geq k^\ell$	$m \geq k^h > k^\ell$ $k^h > n-m-1 \geq k^\ell$
$i \in N_\ell$	$n.s.$	$n.s.$	$n.s.$
$j \in N_h$	$m \geq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d}$	$m \geq \frac{(n-1)(b-d)-k^h(c-d)}{a+b-c-d}$	$m \geq n-1-\frac{a-d}{b-d}k^h$
		$n-m-1 \geq k^h > k^\ell$	
$i \in N_\ell$	$n.s.$	$n.s.$	$n.s.$
$j \in N_h$	$n.s.$	$n.s.$	$n.s.$

Proposition 5.1. *There exist thresholds* $\overline{k^\ell} = \frac{(b-d)(n_\ell-1)-(a-c)n_h}{(c-d)}$ *and* $\underline{k^h} = \frac{(b-d)n_\ell-(a-c)(n_h-1)}{(c-d)}$*, such that:*

i) if $k^\ell < \overline{k^\ell}$ *and* $k^h > \underline{k^h}$*, then* $\mathcal{S}^{\star\star} = \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{BA}$*;*

ii) if $k^\ell \geq \overline{k^\ell}$ *or* $k^h \leq \underline{k^h}$*, then* $\mathcal{S}^{\star\star} = \overrightarrow{AA} \cup \overrightarrow{BB}$*.*

Proposition 5.1 shows that when linking constraints k^ℓ and k^h sufficiently differ from one another, polymorphic states could be contained in $\mathcal{S}^{\star\star}$. This implies that the co-existence of the payoff-dominant action and the risk-dominant action, in the long run, could emerge. Intuitively, agents in the low-constraint group have a constraint low enough such that they can fill sufficiently many of their slots with B-agents. On the other hand, the constraint of agents in the high-constraint group is too large to do so and they will consequently find it optimal to choose A.

As we have seen in above, there may be a multiplicity of absorbing sets under the unperturbed myopic best response learning dynamics. To find which kind of profile is more likely to emerge in the long run we now move forward to characterize which absorbing sets are stochastically stable. In order to do this, we consider a case where agents may make occasional mistakes, which is also known as perturbed myopic best response learning.

Agents are assumed to make mistakes probability $\varepsilon \in (0, 1)$, i.e. they choose a state different to the one prescribed by the unperturbed myopic best response learning dynamics. The probability ε is assumed to be independent across agents, periods, and payoffs. Foster & Young (1990) demonstrate that if the perturbed dynamics is ergodic, irreducible, and aperiodic, then it, which is captured by a Markov process, has a unique invariant distribution $\mu(\epsilon)$ for each fixed ϵ. The limit of this invariant distribution exists and is $\mu^\star = \lim_{\epsilon \to 0} \mu(\epsilon)$. A state s such that $\mu^\star(s) > 0$ is a so-called stochastically stable state or a long-run equilibrium. We denote the set of all stochastically stable states by $\mathcal{S}^{\star\star\star} = \{s \in \mathcal{S} | \mu^\star(s) > 0\}$.

With this technique, we move forward to identify the set of stochastically stable states. Which profile turns out to be stochastically stable will depend on the level of linking constraints. The following propositions establish our main results for various ranges of linking constraints k^ℓ and k^h.

In the first step, we focus on the case where there are only two monomorphic absorbing sets $\overrightarrow{AA}$ and $\overrightarrow{BB}$. After that, we turn to the case where $\overrightarrow{BA}$ is also absorbing.

Proposition 5.2. *For any given $k^h \leq \underline{k^h}$, there exist two thresholds $\underline{\underline{k^\ell}} \leq \overline{\overline{k^\ell}}$, such that: i) if $k^\ell < \underline{\underline{k^\ell}}$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{BB}$; ii) if $k^\ell \in [\underline{\underline{k^\ell}}, \overline{\overline{k^\ell}}]$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{BB} \cup \overrightarrow{AA}$; iii) if $k^\ell > \overline{\overline{k^\ell}}$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{AA}$.*

And for any k^h and k^ℓ such that $k^h > \underline{k^h}$ and $k^\ell \geq \overline{k^\ell}$, we have that $\mathcal{S}^{\star\star\star} = \overrightarrow{AA}$.

Now, we provide technical intuitions for the results by using the case when both constraints are less than half of the number of the other agents, i.e. $k^\ell < k^h < \frac{n-1}{2}$. First, consider the transition from $\overrightarrow{AA}$ to $\overrightarrow{BB}$. Assume that there are k^ℓ agents who mutate to action B and choose any linking strategy. Then, A-agents in N_ℓ will find it optimal to switch to B and link to k^ℓ B-agents. It follows that the number of B-agents now is at least n_ℓ. Since $k^h < \frac{n-1}{2} < n_\ell$, the number of B-agents is sufficient for A-agents in N_h to switch. Thus, k^ℓ mutations are sufficient for this transition, i.e. $r(\overrightarrow{AA}, \overrightarrow{BB}) \leq k^\ell$. Next, consider the transition from $\overrightarrow{BB}$ to $\overrightarrow{AA}$. Note that k_ℓ mutations are insufficient for this transition. To see this, assume that k_ℓ agents mutate to A. After this, there will still be $N - k_\ell$ agents playing B. Since $k^\ell < k^h \leq \frac{n-1}{2}$, we have $k^h \leq n - k^\ell - 1$. This implies that any revising agent (either in N_ℓ or N_h) will find it optimal to either stay with B or switch back to B. It follows that $r(\overrightarrow{BB}, \overrightarrow{AA}) > k^\ell$. We thus have that $r(\overrightarrow{BB}, \overrightarrow{AA}) > r(\overrightarrow{AA}, \overrightarrow{BB})$. Thus, $\overrightarrow{BB}$ is the unique stochastically stable set.

In the Appendix, we provide the proof with both necessary and sufficient conditions for the transitions to occur and thus provide a complete characterization of the set of stochastically stable states for the case where there are only two monomorphic absorbing sets.

Intuitively, when the constraints k^ℓ and k^h are both small, a small number of B-agents is enough for all agents to make choosing the payoff-dominant action optimal. With a logic similar to Staudigl & Weidenholzer (2014), the payoff-dominant convention thus will emerge in the long run. In contrast, if both constraints are sufficiently large, the payoff-dominant action being optimal requires more B-agents to show up. There is increased uncertainty concerning agents' actions with whom one forms links. Consequently, in the long run, agents tend to choose the risk-dominant action, which yields a higher expected payoff.

We now turn to the case where the polymorphic set $\overrightarrow{BA}$ is also absorbing. The following proposition shows our main results of the stochastically stable set when two constraints k^ℓ and k^h are significantly various.

Proposition 5.3. *If $k^h > \underline{k^h}$ and $k^\ell < \overline{k^\ell}$, there exist two thresholds $k^{\ell^\star} < \overline{k^\ell}$ and $k^{h^\star} > \underline{k^h}$, such that whenever $k^\ell \leq k^{\ell^\star}$ and $k^h \geq k^{h^\star}$, $\overrightarrow{BA} \subseteq \mathcal{S}^{\star\star\star}$. Further, for $k^{\ell^{\star\star}} < k^{\ell^\star}$ and $k^{h^{\star\star}} > k^{h^\star}$, such that whenever $k^\ell < k^{\ell^{\star\star}}$ and $k^h > k^{h^{\star\star}}$, $\mathcal{S}^{\star\star\star} = \overrightarrow{BA}$.*

Thus we have identified a region of parameters such that co-existence occurs.[11] Proposition 5.3 shows that if constraints are significantly heterogeneous, the risk-dominant profile and payoff-dominant profile can co-exist. To be more specific, the polymorphic states that agents in the low-

[11]In the cases not covered by the parameter ranges of the Proposition 5.3, i.e. $k^\ell \geq k^{\ell^\star}$ or $k^h \leq k^{h^\star}$, either the risk-dominant convention $\overrightarrow{AA}$ or the payoff-dominant convention $\overrightarrow{BB}$ arises as stochastically stable states. While we have been able to obtain partial results, unfortunately, a complete characterization has eluded us.

constraint group play the payoff-dominant action and agents in the high-constraint group play the risk-dominant action can be stochastically stable if the lower constraint is tighter and the higher constraint is looser.

We now revisit Example 3.2 for the intuition of Proposition 5.3.

Example 3.2 revisited. Recall that when the parameters are $n_\ell = 3, n_h = 2, k^\ell = 1$ and $k^h = 4$, and the payoffs in the coordination game fulfils that $a + 3c \geq 3b + d$, $\overrightarrow{BA}$ is an absorbing set. Figure 4 depicts transitions from monomorphic states to polymorphic states and the other way around, with which we can determine the robustness of these profiles to mistakes. Note that white circles are the agents who play the risk-dominant action B and grey circles are agents who play the payoff-dominant action A.

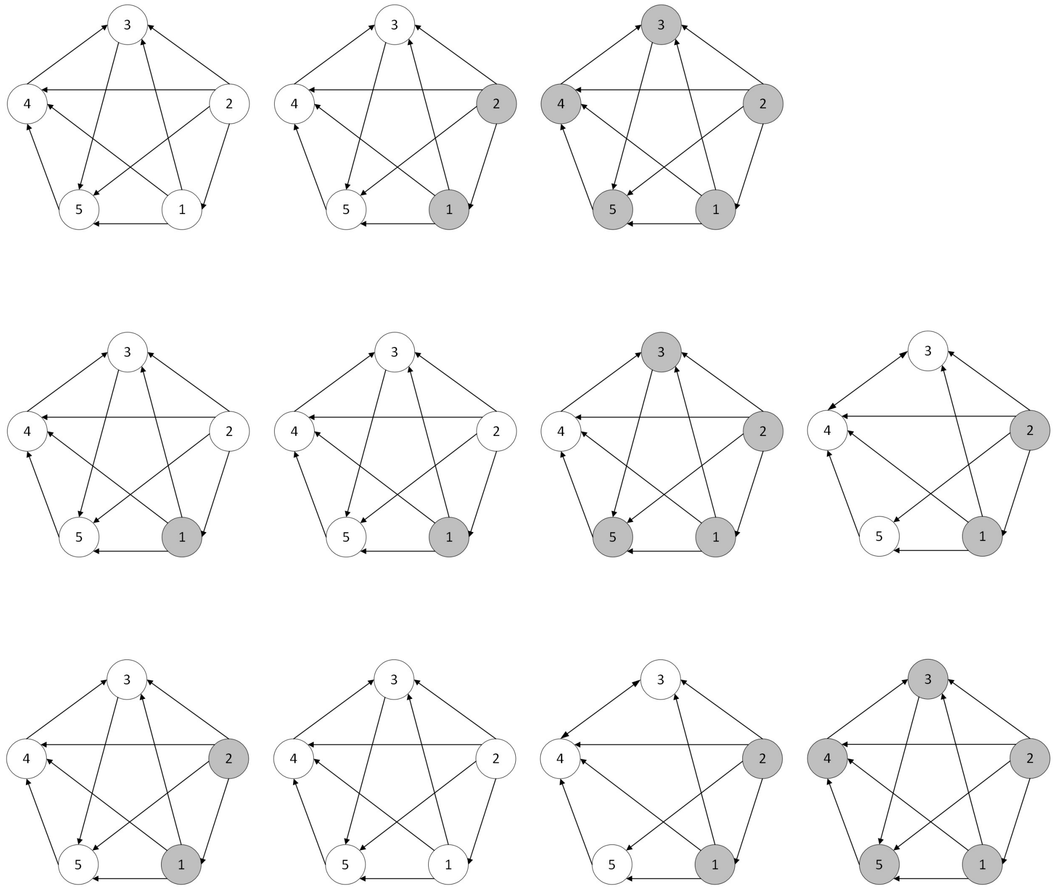
First, we study the transition from $\overrightarrow{BB}$ to $\overrightarrow{BA}$ as Figure 4(a) shows. Agents 1 and 2 can support four links, while agents 3, 4, and 5 can only support one link. Now assume that agent 1 makes a mistake and switches to A. In the next step, agent 2 will also switch since switching to A yields $a + 2c$, which is larger than $3b + d$ from reaming at B. Following this, agents in the low-constraint group will remain at B and link with other B-agents. Thus, with one mistake we have reached a state in $\overrightarrow{BA}$.

Then, consider the transition from $\overrightarrow{BA}$ to $\overrightarrow{BB}$ as Figure 4(b). Assume that agent 2 makes a mistake and switches to B. Following this, agent 1 will switch since there are no other A-agents. B-agents in the low-constraint group will remain since there are sufficient B-agents around. Hence, we have reached a state in $\overrightarrow{BB}$ with one mistake.

The transition from $\overrightarrow{AA}$ to $\overrightarrow{BA}$ is similar to the above (see Figure 4(c)). One mistake is sufficient. To see this, assume that agent 4 makes a mistake and switches to B. In the next step, one B-agent is enough for other agents in the low-constraint group to switch. However, in the present setting, i.e. $a + 3c \geq 3b + d$, agents 1 and 2 may choose to remain. We thus have reached a state in $\overrightarrow{BA}$ with one mutation.

Now, consider the transition from $\overrightarrow{BA}$ to $\overrightarrow{AA}$. For agents in the low-constraint group to switch requires there are no B-agents around. Hence, three mistakes are required for this transition.

Thus, $\overrightarrow{BB}$ and $\overrightarrow{BA}$ can be reached from each absorbing set via a sequence of one mistake for parameters in the present example. Consequently, both $\overrightarrow{BB}$ and $\overrightarrow{BA}$ are stochastically stable. However, $\overrightarrow{AA}$ cannot be reached via such a sequence, implying that more mistakes are required and



(a) $\overrightarrow{BB}$ to $\overrightarrow{BA}$ (b) $\overrightarrow{BA}$ to $\overrightarrow{BB}$ (c) $\overrightarrow{AA}$ to $\overrightarrow{BA}$ (d) $\overrightarrow{BA}$ to $\overrightarrow{AA}$

Figure 4: Tansitions among absorbing sets.

thus $\overrightarrow{AA}$ is not stochastically stable.

Transitions among absorbing sets are similar to the spread of actions. Transitions between two monomorphic absorbing sets $\overrightarrow{AA}$ and $\overrightarrow{BB}$ can be split into two steps: transitions into and out of the intermediate state $\overrightarrow{BA}$. Thus, it can be the case that from $\overrightarrow{AA}$ and $\overrightarrow{BB}$, transitions into $\overrightarrow{BA}$ is easier than out of it. To see this point, if constraints are tight, transition into B is easier, while if constraints are loose, transition into A is easier instead, which is the mechanism that drives the results in Staudigl & Weidenholzer (2014). Note that significantly different constraints in our model imply that the lower constraint is tight and the higher constraint is loose. Thus, for the low-constraint group, the payoff-dominant action B is more robust to mistakes, i.e. transition from $\overrightarrow{AA}$ into $\overrightarrow{BA}$ requires fewer mistakes than the other way around. Similarly, for the high-constraint group, the risk-dominant action A is more robust to mistakes, i.e. transition from $\overrightarrow{BB}$ into $\overrightarrow{BA}$ requires fewer mistakes than the other way around, which is similar to the mechanism in Goyal & Vega-Redondo (2005). Hence, transition into $\overrightarrow{BA}$ from each other may require the fewest mistakes than transitions into the other two absorbing sets $\overrightarrow{AA}$ and $\overrightarrow{BB}$. Consequently, $\overrightarrow{BA}$ can be stochastically stable.

6 Conclusion

In this paper, we present an evolutionary model of coordination and network formation where there are two groups of agents who face either high or low linking constraints on the number of links they can form. We show that the heterogeneous constraints significantly affect the selection of conventions.

The present work reinforces the results of homogeneous constraints where the payoff-dominated action is selected if agents face tight constraints while the risk-dominant action is favoured if the constraints are loose. Moreover, in contrast to the conventional results of only monomorphic states being stochastically stable, we reveal that the co-existence of conventions can be observed when the constraints are significantly different. In this paper, we provide both necessary and sufficient conditions such that the risk-dominant convention and the payoff-dominant convention may co-exist. Specifically, if a large portion of the population has a very tight constraint, with the other having a very loose constraint, the larger group tends to choose the payoff-dominant action, and the smaller group is more likely to choose the risk-dominant action.

Acknowledgments:

I wish to thank Simon Weidenholzer for his supervision and constant help on this paper. I am also indebted to my defence committees Ennio Bilancini and Albin Erlanson, for their helpful comments.

Financial Support:

I acknowledge financial support from the University of Essex Social Sciences Doctoral Scholarship and the Swiss National Science Foundation (SNSF).

References

Albert, R., Jeong, H. & Barabási, A.-L. (1999), 'Diameter of the world-wide web', Nature **401**(6749), 130–131.

Anwar, A. W. et al. (2002), 'On the co-existence of conventions', Journal of Economic Theory **107**(1), 145–155.

Bhaskar, V. & Vega-Redondo, F. (2004), 'Migration and the evolution of conventions', Journal of Economic Behavior & Organization **55**(3), 397–418.

Bilancini, E. & Boncinelli, L. (2018), 'Social coordination with locally observable types', Economic Theory **65**(4), 975–1009.

Blume, L. E. (1993), 'The statistical mechanics of strategic interaction', Games and Economic Behavior **5**(3), 387–424.

Blume, L. E. (1995), 'The statistical mechanics of best-response strategy revision', Games and Economic Behavior **11**(2), 111–145.

Cui, Z. & Shi, F. (2022), 'Bandwagon effects and constrained network formation', Games and Economic Behavior **134**, 37–51.

Cui, Z. & Weidenholzer, S. (2021), 'Lock-in through passive connections', Journal of Economic Theory **192**, 105187.

Dieckmann, T. (1999), 'The evolution of conventions with mobile players', Journal of Economic Behavior & Organization **38**(1), 93–111.

Dutta, B. & Mutuswami, S. (1997), 'Stable networks', Journal of Economic Theory **76**(2), 322–344.

Ellison, G. (1993), 'Learning, local interaction, and coordination', Econometrica pp. 1047–1071.

Ellison, G. (2000), 'Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution', The Review of Economic Studies **67**(1), 17–45.

Foster, D. & Young, P. (1990), 'Stochastic evolutionary game dynamics', Theoretical Population Biology **38**(2), 219–232.

Freidlin, M. I. & Wentzell (1998), Random perturbations, Springer.

Goyal, S., Van Der Leij, M. J. & Moraga-González, J. L. (2006), 'Economics: An emerging small world', Journal of Political Economy **114**(2), 403–412.

Goyal, S. & Vega-Redondo, F. (2005), 'Network formation and social coordination', Games and Economic Behavior **50**(2), 178–207.

Harsanyi, J. C. & Selten, R. (1988), 'A general theory of equilibrium selection in games', MIT Press Books **1**.

Jackson, M. O. & Rogers, B. W. (2007), 'Meeting strangers and friends of friends: How random are social networks?', American Economic Review **97**(3), 890–915.

Jackson, M. O. & Watts, A. (2002), 'On the formation of interaction networks in social coordination games', Games and Economic Behavior **41**(2), 265–291.

Jackson, M. O. & Wolinsky, A. (1996), 'A strategic model of social and economic networks', Journal of Economic Theory **71**, 44–74.

Kandori, M., Mailath, G. J. & Rob, R. (1993), 'Learning, mutation, and long run equilibria in games', Econometrica pp. 29–56.

Kwak, H., Lee, C., Park, H. & Moon, S. (2010), What is twitter, a social network or a news media?, Vol. 19.

Lu, F. & Shi, F. (2023), 'Coordination with heterogeneous interaction constraints', Games and Economic Behavior **142**, 645–665.

Oechssler, J. (1997), 'Decentralization and the coordination problem', Journal of Economic Behavior & Organization **32**(1), 119–135.

Pin, P., Weidenholzer, E. & Weidenholzer, S. (2017), 'Constrained mobility and the evolution of efficient outcomes', Journal of Economic Dynamics and Control **82**, 165–175.

Staudigl, M. & Weidenholzer, S. (2014), 'Constrained interactions and social coordination', Journal of Economic Theory **152**, 41–63.

Young, H. P. (1993), 'The evolution of conventions', Econometrica pp. 57–84.

Zeng, Q. (2019), 'Social coordination and network formation with heterogeneous constraints', Game Theory for Networking Applications pp. 153–162.

A Appendix

A.1 Proofs of Section 3

Proof of Lemma 1. First, consider the case where some agents do not form the maximum number of links they can support. Given $\gamma < d < c < a < b$, those agents will fill their remaining slots as any extra link yields at least $d - \gamma$. As a result, every agent will form the maximum number of links.

Next, consider the case where agents within the same group play different actions, e.g. some agents in the low-constraint group play action A and the other agents play action B. Note that the LOPs for any agent in the same group are identical given any strategy profile. First, consider agents in the low-constraint group. Their LOPs are given by

$$v(A, m) = a \cdot \min\{k^\ell, m-1\} + c \cdot (k^\ell - \min\{k^\ell, m-1\}) - \gamma \cdot k^\ell \quad (3)$$

and

$$v(B, m) = b \cdot \min\{k^\ell, n-m-1\} + d \cdot (k^\ell - \min\{k^\ell, n-m-1\}) - \gamma \cdot k^\ell \quad (4)$$

Consider two agents i and j in N_ℓ. Assume that agent i plays action A and agent j plays action B. If agent i behaves optimally, then it must be the case $v(A, m) \geq v(B, m-1)$. By equations (3) and (4), we have

$$a \cdot \min\{k^\ell, m-1\} + c \cdot (k^\ell - \min\{k^\ell, m-1\}) \geq b \cdot \min\{k^\ell, n-m\} + d \cdot (k^\ell - \min\{k^\ell, n-m\}). \quad (5)$$

Similarly, agent j behaves optimally if $v(B,m) \geq v(A,m+1)$. We thus have

$$b \cdot \min\{k^\ell, n-m-1\} + d \cdot (k^\ell - \min\{k^\ell, n-m-1\}) \geq a \cdot \min\{k^\ell, m\} + c \cdot (k^\ell - \min\{k^\ell, m\}). \quad (6)$$

Note that for agents in the low-constraint group to choose different actions as best responses, inequalities (5) and (6) have to hold simultaneously. We solve inequalities (5) and (6) independently by discussing different levels of k^ℓ. Table 2 presents the solutions for these two inequalities for various relevant ranges of thresholds.

One can check that inequalities (5) and (6) never have a common solution for various levels of k^ℓ. This implies that both agents can't behave optimally simultaneously. Thus, a strategy profile where agents in the low-constraint group play different actions is not a Nash equilibrium.

The argument for agents in the high-constraint group follows the same logic and is omitted. Therefore, for any $s \notin \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{AB} \cup \overrightarrow{BA}$, s is not a Nash equilibrium. □

Table 2: Conditions on m such that i and j behave optimally at different levels of k^ℓ.

	$k^\ell > m-1$ and $k^\ell > n-m$	$k^\ell \leq m-1$ and $k^\ell > n-m$	$k^\ell \leq n-m$
$v(A,m) \geq v(B,m-1)$	$m \geq \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d} + 1$	$m \geq n - \frac{a-d}{b-d}k^\ell$	never
	$k^\ell > m$ and $k^\ell > n-m-1$	$k^\ell \leq m$ and $k^\ell > n-m-1$	$k^\ell \leq n-m-1$
$v(B,m) \geq v(A,m+1)$	$m \leq \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d}$	$m \leq n-1 - \frac{a-d}{b-d}k^\ell$	always

Proof of Lemma 3. The proof proceeds by contradiction. Consider two agents $i \in N_\ell$ and $j \in N_h$. Note that the distribution of actions in a strategy profile $s \in \overrightarrow{AB}$ is (n_ℓ, n_h). Then agent i is now playing action A and agent j is playing action B. We assume that s is a Nash equilibrium, then neither i nor j will deviate.

Consider agent i. Agent i will stay with A if

$$\begin{aligned} v_i(A, n_\ell) &= a \cdot \min\{k^\ell, n_\ell - 1\} + c \cdot (k^\ell - \min\{k^\ell, n_\ell - 1\}) - M \cdot k^\ell \\ &\geq b \cdot \min\{k^\ell, n_h\} + d \cdot (k^\ell - \min\{k^\ell, n_h\}) - M \cdot k^\ell = v_i(B, n_\ell - 1). \end{aligned} \quad (7)$$

Similarly, agent j will stay with B if

$$\begin{aligned} v_j(B, n_\ell) &= b \cdot \min\{k^h, n_h - 1\} + d \cdot (k^h - \min\{k^h, n_h - 1\}) - M \cdot k^h \\ &\geq a \cdot \min\{k^h, n_\ell\} + c \cdot (k^h - \min\{k^h, n_\ell\}) - M \cdot k^h = v_j(A, n_\ell + 1). \end{aligned} \tag{8}$$

Note that in a Nash equilibrium, inequalities (7) and (8) have to hold simultaneously. First, we solve inequality (7) to find the switching threshold for agent i. We have three sub-cases by considering the order of k^ℓ, $n_\ell - 1$, and n_h.

i) if $k^\ell \leq n_h \leq n_\ell - 1$, we have

$$a \cdot k^\ell + c \cdot (k^\ell - k^\ell) \geq b \cdot k^\ell + d \cdot (k^\ell - k^\ell) \Rightarrow a \cdot k^\ell \geq b \cdot k^\ell \Rightarrow a \geq b$$

which is in contradiction to the order of payoffs in the coordination game. Thus, agent i will deviate.

ii) if $n_h < k^\ell \leq n_\ell - 1$, we have

$$a \cdot k^\ell + c \cdot (k^\ell - k^\ell) \geq b \cdot n_h + d \cdot (k^\ell - n_h) \Rightarrow k^\ell \geq \frac{b - d}{a - d} \cdot n_h. \tag{9}$$

One can check that inequality (9) holds iff $\frac{b-d}{a-d} \cdot n_h \leq n_\ell - 1$.[12] Note that $\frac{b-d}{a-d} \cdot n_h > n_h$ holds as $b-d > a-d$. The solution to inequality (9) is $\frac{b-d}{a-d} \cdot n_h \leq k^\ell \leq n_\ell - 1$. Moreover, if $\frac{b-d}{a-d} \cdot n_h \leq n_\ell - 1$, then we have that $n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d)$. Thus, agent i will stay with A if $n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d)$.

iii) if $n_h \leq n_\ell - 1 < k^\ell$, we have

$$a \cdot (n_\ell - 1) + c \cdot (k^\ell - (n_\ell - 1)) \geq b \cdot n_h + d \cdot (k^\ell - n_h) \Rightarrow k^\ell \geq \frac{(b - d)n_h - (a - c)(n_\ell - 1)}{c - d}.$$

Agent i will stay with A if $k^\ell \geq \max\{\frac{(b-d)n_h - (a-c)(n_\ell - 1)}{c-d}, n_\ell - 1\}$. Furthermore, one can check that $n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d)$ if $\frac{(b-d)n_h - (a-c)(n_\ell - 1)}{c-d} \leq n_\ell - 1$. Thus, we have that $k^\ell > n_\ell - 1$ if $n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d)$ and $k^\ell \geq \frac{(b-d)n_h - (a-c)(n_\ell - 1)}{c-d}$ if $n_h > \frac{a-d}{b-d} \cdot n_\ell - (a-d)$.

[12] If $\frac{b-d}{a-d} \cdot n_h > n_\ell - 1$, then $n_h > \frac{a-d}{b-d} \cdot n_\ell - (a-d)$. Thus, we have that $k^\ell \leq n_\ell - 1$ which contradicts $k^\ell \geq \frac{b-d}{a-d} \cdot n_h$. In this case, inequality (9) does not hold. This implies that agent i will deviate.

Summary up, agent i will stay with A if

$$k^\ell \geq \begin{cases} \frac{b-d}{a-d} \cdot n_h, & \text{if } n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d). \\ \frac{(b-d)n_h-(a-c)(n_\ell-1)}{c-d}, & \text{if } n_h > \frac{a-d}{b-d} \cdot n_\ell - (a-d). \end{cases} \tag{10}$$

Similarly, we solve inequality (8) by considering various ranges of k^h.

i) if $k^h \leq n_h - 1 < n_\ell$, we have

$$b \cdot k^h + d \cdot (k^h - k^h) \geq a \cdot k^h + c \cdot (k^h - k^h) \Rightarrow b \cdot k^h \geq a \cdot k^h \Rightarrow b \geq a$$

which is consistent with the order of payoffs in the coordination game. Thus, agent j will stay with B.

ii) if $n_h - 1 < k^h < n_\ell$, we have

$$b \cdot (n_h - 1) + d \cdot (k^h - (n_h - 1)) \geq a \cdot k^h + c \cdot (k^h - k^h) \Rightarrow k^h \leq \frac{b-d}{a-d} \cdot (n_h - 1).$$

Agent j will stay with B if $k^h \leq \min\{\frac{b-d}{a-d} \cdot (n_h-1), n_\ell\}$. Moreover, we obtain that $n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$ from $\frac{b-d}{a-d} \cdot (n_h - 1) < n_\ell$. Thus, we have that $k^h \leq \frac{b-d}{a-d} \cdot (n_h-1)$ if $n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$ and $k^h < n_\ell$ if $n_h \geq \frac{a-d}{b-d} \cdot n_\ell + (b-d)$.

iii) if $n_h - 1 < n_\ell \leq k^h$, we have

$$b \cdot (n_h - 1) + d \cdot (k^h - (n_h - 1)) \geq a \cdot n_\ell + c \cdot (k^h - n_\ell) \Rightarrow k^h \leq \frac{(b-d)(n_h-1) - (a-c)n_\ell}{c-d}. \tag{11}$$

Inequality (11) has solution if and only if $\frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d} \geq n_\ell$.[13] Furthermore, we have that $n_h \geq \frac{a-d}{b-d} \cdot n_\ell + (b-d)$ if $\frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d} \geq n_\ell$. Thus, agent j will stay with B if $n_\ell \leq k^h \leq \frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d}$, in the case where $n_h \geq \frac{a-d}{b-d} \cdot n_\ell + (b-d)$.

[13] If $\frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d} < n_\ell$, then $n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$. Thus, we have that $k^h \leq \frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d}$ which contradicts $k^h \geq n_\ell$. In this case, inequality (11) does not have a solution, and agent j will deviate.

Summary up, agent j will stay with B if

$$k^h \leq \begin{cases} \frac{b-d}{a-d} \cdot (n_h - 1), & \text{if } n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d). \\ \frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d}, & \text{if } n_h \geq \frac{a-d}{b-d} \cdot n_\ell + (b-d). \end{cases} \tag{12}$$

Note that equations (10) and (12) have to hold simultaneously in a Nash equilibrium. To find the solution to these two equations, there are three sub-cases for the various ranges of n_ℓ and N_h.

First, consider the sub-case where $n_h \leq \frac{a-d}{b-d} \cdot n_\ell - (a-d)$. We have that agent i will stay with A if $k^\ell \geq \frac{b-d}{a-d} \cdot n_h$. As $\frac{a-d}{b-d} \cdot n_\ell - (a-d) < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$, we have agent j will stay with B if and only if $k^h \leq \frac{b-d}{a-d} \cdot (n_h - 1)$. Therefore, the condition for both i and j staying in their action is $k^h \leq \frac{b-d}{a-d} \cdot (n_h - 1) < \frac{b-d}{a-d} \cdot n_h \leq k^\ell$, which contradicts our assumption $k^\ell < k^h$. Thus, either agent i or j will switch.

Next, consider the sub-case where $\frac{a-d}{b-d} \cdot n_\ell - (a-d) < n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$. We have that agent i will stay with A if $k^\ell \leq \frac{(b-d)n_h-(a-c)(n_\ell-1)}{c-d}$ and agent j will stay with B if $k^h \leq \frac{b-d}{a-d} \cdot (n_h - 1)$. Since $n_h > \frac{a-d}{b-d} \cdot n_\ell - (a-d)$ and $k^\ell \geq \frac{(b-d)n_h-(a-c)(n_\ell-1)}{c-d}$, we obtain that $k^\ell > n_\ell - 1$. And since $n_h < \frac{a-d}{b-d} \cdot n_\ell + (b-d)$ and $k^h \leq \frac{b-d}{a-d} \cdot (n_h - 1)$, we have that $k^h < n_\ell$. Moreover, since n_ℓ is an integer, $k^\ell \geq n_\ell - 1$ implies that $k^\ell \geq n_\ell$, and $k^h < n_\ell$ implies that $k^h \leq n_\ell - 1$. Thus, we have that $k^h \leq n_\ell - 1 < n_\ell \leq k^\ell$, which contradicts $k^\ell < k^h$. Therefore, either agent i or j will deviate.

Finally, consider the sub-case where $n_h \geq \frac{a-d}{b-d} \cdot n_\ell + (b-d)$. We have that agent j will stay with B if $k^h \leq \frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d}$. As $\frac{a-d}{b-d} \cdot n_\ell + (b-d) > \frac{a-d}{b-d} \cdot n_\ell - (a-d)$, from equation (10) we find that agent i will stay with A if and only if $k^\ell \geq \frac{(b-d)n_h-(a-c)(n_\ell-1)}{c-d}$. Thus, the condition for both agents staying in their actions is $k^\ell \geq \frac{(b-d)n_h-(a-c)(n_\ell-1)}{c-d} > \frac{(b-d)(n_h-1)-(a-c)n_\ell}{c-d} \geq k^h$, which contradicts $k^\ell < k^h$. Therefore, either agent i or j will switch in this sub-case.

Consequently, there does not exist any n_ℓ and n_h such that both agents i and j stay with their actions, i.e. either i or j will deviate. Thus, $s \in \overrightarrow{AB}$ is not a Nash equilibrium. $\square$

Proof of Lemma 4. Note that the distribution of actions in a strategy profile $s \in \overrightarrow{BA}$ is (n_h, n_ℓ). Consider two agents $i \in N_\ell$ and $j \in N_h$. Note that agent i is playing action B and agent j is playing action A. As s is a Nash equilibrium, neither i nor j will deviate from their current actions.

First, consider agent i. She will stay with B if and only if

$$\begin{aligned} v_i(B,n_h) &= b\cdot\min\{k^\ell,n_\ell-1\}+d\cdot(k^\ell-\min\{k^\ell,n_\ell-1\})-M\cdot k^\ell \\ &\geq a\cdot\min\{k^\ell,n_h\}+c\cdot(k^\ell-\min\{k^\ell,n_h\})-M\cdot k^\ell = v_i(A,n_h+1). \end{aligned} \tag{13}$$

Similarly, agent j will stay with A if and only if

$$\begin{aligned} v_j(A,n_h) &= a\cdot\min\{k^h,n_h-1\}+c\cdot(k^h-\min\{k^h,n_h-1\})-M\cdot k^h \\ &\geq b\cdot\min\{k^h,n_\ell\}+d\cdot(k^h-\min\{k^h,n_\ell\})-M\cdot k^h = v_j(B,n_h-1). \end{aligned} \tag{14}$$

Note that in a Nash equilibrium, inequalities (13) and (14) have to hold simultaneously. First, we solve inequalities (13) to obtain switching thresholds for agent i. There are three sub-cases by considering various orders of k^ℓ, $n_\ell-1$, and n_h.

i) if $k^\ell\leq n_h<n_\ell-1$, we have

$$b\cdot k^\ell+d\cdot(k^\ell-k^\ell)\geq a\cdot k^\ell+c\cdot(k^\ell-k^\ell)\Rightarrow b\cdot k^\ell\geq a\cdot k^\ell\Rightarrow b\geq a$$

which coincides with the order of payoffs in the coordination game. Thus, agent i will stay.

ii) if $n_h<k^\ell\leq n_\ell-1$, we have

$$b\cdot k^\ell+d\cdot(k^\ell-k^\ell)\geq a\cdot n_h+c\cdot(k^\ell-n_h)\Rightarrow k^\ell\geq\frac{a-c}{b-c}\cdot n_h.$$

Note that $\frac{a-c}{b-c}\cdot n_h<n_h$ holds as $a-c<b-c$. Following that, we have $k^\ell>\frac{a-c}{b-c}\cdot n_h$ whenever $n_h<k^\ell\leq n_\ell-1$. Thus, inequality (13) holds in the relevant range of k^ℓ. This implies that agent i will stay with action B whenever $n_h<k^\ell\leq n_\ell-1$.

iii) if $n_h\leq n_\ell-1<k^\ell$, we have

$$b\cdot(n_\ell-1)+d\cdot(k^\ell-(n_\ell-1))\geq a\cdot n_h+c\cdot(k^\ell-n_h)\Rightarrow k^\ell\leq\frac{(b-d)(n_\ell-1)-(a-c)n_h}{c-d}=\overline{k^\ell}.$$

One can check that $\overline{k^\ell}>(n_\ell-1)$.[14] Thus, inequality (13) holds if and only if $n_\ell-1<k^\ell\leq\overline{k^\ell}$

[14] This is obtained by considering $\frac{(b-d)(n_\ell-1)-(a-c)n_h}{c-d}-(n_\ell-1)=\frac{(b-d)(n_\ell-1)-(a-c)n_h-(c-d)(n_\ell-1)}{c-d}=\frac{(b-c)(n_\ell-1)-(a-c)n_h}{c-d}\geq\frac{(b-c)n_h-(a-c)n_h}{c-d}>0$.

and agent i will stay with B in the relevant range. As we have seen in cases i) and ii), agents will stay with B if $k^\ell \leq n_h < n_\ell - 1$ and $n_h < k^\ell \leq n_\ell - 1$. Combining these results with the condition obtained in case iii) yields that agent i will stay with action B if and only if $k^\ell \leq \overline{k^\ell}$.

Now, consider agent j. Similarly, we solve inequality (14) by considering different orders of k^h, $n_h - 1$ and n_ℓ.

i) if $k^h \leq n_h - 1 < n_\ell$, we have

$$a \cdot k^h + c \cdot (k^h - k^h) \geq b \cdot k^h + d \cdot (k^h - k^h) \Rightarrow a \cdot k^h \geq b \cdot k^h \Rightarrow a \geq b$$

which contradicts the order of payoffs $b > a$. Thus, agent j will deviate.

ii) if $n_h - 1 < k^h < n_\ell$, we have

$$a \cdot (n_h - 1) + c \cdot (k^h - (n_h - 1)) \geq b \cdot k^h + d \cdot (k^h - k^h) \Rightarrow k^h \leq \frac{a-c}{b-c} \cdot (n_h - 1).$$

Note that $\frac{a-c}{b-c} \cdot (n_h - 1) < n_h - 1$ holds since $b > a$. There is a contradiction between $k^h > n_h - 1$ and $k^h \leq \frac{a-c}{b-c} \cdot (n_h - 1)$. This implies that inequality (14) does not hold and agent j will deviate in this range.

iii) if $n_h - 1 < n_\ell \leq k^h$, we have

$$a \cdot (n_h - 1) + c \cdot (k^h - (n_h - 1)) \geq b \cdot n_\ell + d \cdot (k^h - n_\ell) \Rightarrow k^h \geq \frac{(b-d)n_\ell - (a-c)(n_h - 1)}{c-d} = \underline{k^h}.$$

Note that $\underline{k^h} > n_\ell$.[15] Thus, agent j will stay with action A if and only if $k^h \geq \underline{k^h}$. Attending the results in cases i) and ii), agents will stay with B if $k^h \leq n_h - 1 < n_\ell$ and $n_h - 1 < k^h \leq n_\ell$. Combining these results with the condition obtained in case iii) yields that agent j will stay with action A if and only if $k^h \geq \underline{k^h}$.

Consequently, for both agents i and j to stay with their current actions requires that $k^\ell \leq \overline{k^\ell}$ and $k^h \geq \underline{k^h}$. Furthermore, if both $k^\ell \leq \overline{k^\ell}$ and $k^h \geq \underline{k^h}$ hold, then both agents i and j with stay with their current actions. $\square$

[15]This is obtained by considering $\frac{(b-d)n_\ell - (a-c)(n_h-1)}{c-d} - n_\ell = \frac{(b-d)n_\ell - (a-c)(n_h-1) - (c-d)n_\ell}{c-d} = \frac{(b-c)n_\ell - (a-c)(n_h-1)}{c-d} > \frac{(b-c)n_h - (a-c)n_h + (a-c)}{c-d} > 0$.

A.2 Proofs of Section 4

Proof of Proposition 5.1. This proof proceeds in two steps. In the first step, we prove that from any state s, the dynamics leads to a monomorphic or polymorphic state, i.e. a state $s' \in \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{AB} \cup \overrightarrow{BA}$, with a positive probability. In the second step, we show that from each state s', this process converges to a Nash equilibrium.

We prove the first step by constructing a sequence of revisions leading to a monomorphic or polymorphic state from any state s. This sequence of revisions consists of multiple rounds where in each round, one of the two groups is selected and all agents in this selected group can revise their strategies.[16] Moreover, we assume that if agents are indifferent between two actions, they will remain with their current actions.[17]

Consider an initial state s with distribution of actions $(m, n-m)$. In the first round, give the revision opportunity to agents in N_ℓ. Consider the case where A-agents in N_ℓ remain. This implies that $v(A,m) \geq v(B,m-1)$, i.e. $m \geq M_\ell^1$ (or $m \geq M_\ell^2$). Table 1 reveals that for any B-agents in N_ℓ, the optimal choice is to switch. Thus, we have reached a state where all agents in N_ℓ play action A. Now, consider the case where A-agents in N_ℓ switch. This implies that $v(B,m-1) > v(A,m)$, i.e. $m < M_\ell^1$ (or $m < M_\ell^2$). This implies that $m \leq M_\ell^1 - 1$ (or $m \leq M_\ell^2 - 1$), and furthermore, $v(A,m+1) \leq v(B,m)$. Thus, B-agents will remain and we have reached a state where all agents in N_ℓ play action B.

In the second round, give the revision opportunity to agents in N_h. Assume that the distribution of actions is $(m', n-m')$ after agents in N_ℓ have revised. Agents in N_h decide whether to remain or to switch based on this new distribution of actions. Following similar arguments as for those agents in N_ℓ, we will arrive at a state where all agents in N_h play the same action A or B.

Consequently, after two rounds of revisions, we have reached a state where agents in the same group play the same action, i.e. a state $s' \in \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{AB} \cup \overrightarrow{BA}$.

In the second step, we show this process will converge to a Nash equilibrium from any state s'. First, consider any state $s' \in \overrightarrow{AA} \cup \overrightarrow{BB}$. If all agents play action A (and also B), no one will switch since $v_i(A,n) > v_i(B,n-1), \forall i \in N$ (and since $v_i(B,0) > v_i(A,1)$).

[16] This sequence occurs with positive probability since the probability of each agent receiving the revision opportunity is positive.

[17] Note that since agents randomize between two actions when they are indifferent, the probability of their remains is positive.

Now consider a state $s' \in \overrightarrow{AB}$. If agents in N_ℓ find it optimal to play action A, then we have that $v_i(A, n_\ell) > v_i(B, n_\ell - 1), \forall i \in N_\ell$, which implies that $n_\ell \geq M_\ell^1$ (or $n_\ell \geq M_\ell^2$). Note that $M_\ell^2 > M_h^2$ and $M_\ell^1 > M_h^1$ holds (see in Table 1). We thus obtain that $n_\ell > M_h^1$ (or $n_\ell > M_h^2$). Table 1 reveals that it is optimal for B-agents in N_h to switch to A. Similarly, one can check that if agents in N_h find it optimal to play action B, then it is optimal for A-agents in N_ℓ to switch to B. Thus, we will arrive at a state $s \in \overrightarrow{AA} \cup \overrightarrow{BB}$.

Then, consider a state $s' \in \overrightarrow{BA}$. In the proof of Lemma 4 we have argued that it is optimal for agents in N_ℓ to play B and for agents in N_h to play A iff $k^\ell \leq \overline{k^\ell}$ and $k^h \geq \underline{k^h}$. This implies that whenever $k^\ell < \overline{k^\ell}$ and $k^h > \underline{k^h}$, agents will strictly prefer to remain at their actions when they receive the revision opportunity. It follows that if $k^\ell \geq \overline{k^\ell}$, agents in N_ℓ will find it optimal to play A and switch. Similarly, if $k^h \leq \underline{k^h}$, agents in N_h will find it optimal to play B and switch. Then we will reach a state $s \in \overrightarrow{AA} \cup \overrightarrow{BB}$.

Consequently, this process will finally converge to a state $s \in \overrightarrow{AA} \cup \overrightarrow{BB} \cup \overrightarrow{BA}$ if $k^\ell < \overline{k^\ell}$ and $k^h > \underline{k^h}$, and will converge to a state $s \in \overrightarrow{AA} \cup \overrightarrow{BB}$ otherwise. According to Proposition 4.1, s is a Nash equilibrium for the relevant ranges of k^ℓ and k^h.

Now, we proceed to show that this process moves between any pair of states s and s' in $\overrightarrow{AA}$ (and also for any pair in $\overrightarrow{BB}$ and $\overrightarrow{BA}$) with positive probability. Note that s and s' only differ in the linking strategies of agents. As agents are indifferent between linking to any of those agents with the same action, this process will move between any such two strategies with a positive probability. Thus, all states in $\overrightarrow{AA}$ (also in $\overrightarrow{BB}$ and $\overrightarrow{BA}$) form an absorbing set.[18] $\square$

A.3 Proofs of Section 5

Proof of Proposition 5.2. First, note that if $k^\ell \geq \underline{k^\ell}$ or $k^h \leq \overline{k^h}$, $\overrightarrow{AA}$ and $\overrightarrow{BB}$ are the only two absorbing sets.

First, consider the transition from $\overrightarrow{AA}$ to $\overrightarrow{BB}$.

Note that A-agents with the lower constraint k^ℓ require fewer mutations to switch than agents with the higher constraint k^h. To find the minimum number of mutations required for this transition, we thus start with agents in N_ℓ. Denote by $n - m_\ell^{AB}$ the minimum number of B-agents required for

[18] The discussion regarding states in $\overrightarrow{BA}$ is consistently established on the condition $k^\ell < \overline{k^\ell}$ and $k^h > \underline{k^h}$.

the successful transition of agents in N_ℓ. Consequently, m_ℓ^{AB} is the maximum number of remaining A-agents.

First, note that A-agents with the lower constraint will always switch if $m_\ell^{AB} \leq n - k^\ell$. Since that $m_\ell^{AB} \leq n - k^\ell$ implies $k^\ell \leq n - m_\ell^{AB}$, the number of B-agents is sufficient such that A-agents can fill all their slots with B-agents. Thus, we now turn to the case where $m_\ell^{AB} > n - k^\ell$, i.e. the number of B-agents is insufficient such that A-agents cannot fill all their slots with B-agents. We now need to determine the payoff A-agents get when they stay with A. Hence we need to distinguish two sub-cases: i) A-agents can fill all their slots with other A-agents, i.e. $m_\ell^{AB} \geq k^\ell + 1$, and ii) A-agents have to link to both A- and B-agents, i.e. $m_\ell^{AB} < k^\ell + 1$.

Consider sub-case i). According to Table 1, the switching threshold for A-agents is $m_\ell^{AB} = \lfloor M_\ell^2 \rfloor$. Observe now that this sub-case will happen if indeed $m_\ell^{AB} = \lfloor M_\ell^2 \rfloor \geq k^\ell + 1$. Attending to the definition of M_ℓ^2 and solving for k^ℓ, we have that $k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}$.

Now, consider sub-case ii). According to Table 1, the switching threshold for A-agents in this sub-case is given by $m_\ell^{AB} = \lfloor M_\ell^1 \rfloor$. Solving for k^ℓ yields $k^\ell > \frac{(n-1)(b-d)}{a+b-2d}$.

Recall that if there are $n - k^\ell$ or less A-agents, all agents in N_ℓ will switch to B. Thus, the maximum number of A-agents for the transition to occur is characterized by

$$m_\ell^{AB} = \begin{cases} \max\{\lfloor M_\ell^2 \rfloor, n - k^\ell\}, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ \max\{\lfloor M_\ell^1 \rfloor, n - k^\ell\}, & \text{if } k^\ell > \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

One can check that $\lfloor M_\ell^2 \rfloor > n - k^\ell$ always holds for any k^ℓ, and $\lfloor M_\ell^1 \rfloor > n - k^\ell$ holds whenever $k^\ell > \frac{(n-1)(a-c)}{a+b-2c}$. Since $\frac{(n-1)(a-c)}{a+b-2c} < \frac{(n-1)(b-d)}{a+b-2d}$, we thus have

$$m_\ell^{AB} = \begin{cases} \lfloor M_\ell^2 \rfloor, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ \lfloor M_\ell^1 \rfloor, & \text{if } k^\ell > \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

Therefore, the minimum number of mutations required for the transition of A-agents in N_ℓ is

$$n - m_\ell^{AB} = \begin{cases} n - \lfloor M_\ell^2 \rfloor, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^1 \rfloor, & \text{if } k^\ell > \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

Now, we assess the largest number of B-agents after the mutations and switches (excluding switches among agents in N_h for now), i.e. agents who have mutated and agents in N_ℓ who have switched to B. For this, assume all mutations occur in N_h. Thus, the number of B-agents is $n_\ell + n - m_\ell^{AB}$. It follows that the number of remaining A-agents in N_h is $m_\ell^{AB} - n_\ell$. We have that $m_\ell^{AB} - n_\ell < n_h$ holds for any relevant range of k^ℓ since the mutations occur among N_h and every agent in N_ℓ switched.

Now consider agents in N_h. First, denote by m_h^{AB} the number of A-agents required for agents in N_h to switch. Following the same argument as for agents in N_ℓ, attending to Table 1 reveals that m_h^{AB} is given by

$$m_h^{AB} = \begin{cases} \lfloor M_h^2 \rfloor, & \text{if } k^h \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ \lfloor M_h^1 \rfloor, & \text{if } k^h > \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

If the number of A-agents $m_\ell^{AB} - n_\ell$ is less than m_h^{AB}, then agents in N_h switch without requiring more mutations. Otherwise, extra mutations are needed for their transition.

First, consider the case where $k^h \leq \frac{(n-1)(b-d)}{a+b-2d}$. One can check that $\lfloor M_h^2 \rfloor > n_h$ holds. Thus, we have that $m_\ell^{AB} - n_\ell < \lfloor M_h^2 \rfloor$, i.e. the number of existing A-agents is smaller than the number of A-agents required for the transition. Therefore, the number of B-agents is sufficient for A-agents in N_h to switch.

Next, consider the case where $\frac{(n-1)(b-d)}{a+b-2d} < k^h \leq \underline{k^h}$. We have that $\lfloor M_h^1 \rfloor \geq n_h$ holds. This implies that no extra mutation is required for A-agents in N_h to switch.

Now, consider the case where $k^h > \underline{k^h}$. We have that $\lfloor M_h^1 \rfloor < n_h$. Recall that we are now focusing on the case where there are only two absorbing sets. Thus, the range of k^ℓ is restricted on $k^\ell \geq \overline{k^\ell}$ whenever $k^h > \underline{k^h}$. One can check that $m_\ell^{AB} - n_\ell \leq 0$ if $k^\ell \geq \overline{k^\ell}$. Note that if $m_\ell^{AB} - n_\ell \leq 0$, all agents are now playing B and we have reached $\overrightarrow{BB}$.

Combining the results of all three cases, the number of B-agents after mutations and switches among N_ℓ is sufficient for agents in N_h to switch. Denote by $n - m^{AB}$ the minimum number of

B-agents for the transition among N_ℓ and N_h. In summary, we have that

$$n - m^{AB} = \begin{cases} n - \lfloor M_\ell^2 \rfloor, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}, k^h \leq \frac{(b-d)n_\ell-(a-c)(n_h-1)}{(c-d)}. \\ n - \lfloor M_\ell^1 \rfloor, & \text{if } k^\ell > \frac{(n-1)(b-d)}{a+b-2d}, k^h \leq \frac{(b-d)n_\ell-(a-c)(n_h-1)}{(c-d)}. \\ n - \lfloor M_\ell^1 \rfloor, & \text{if } k^\ell \geq \frac{(b-d)(n_\ell-1)-(a-c)n_h}{(c-d)}, k^h > \frac{(b-d)n_\ell-(a-c)(n_h-1)}{(c-d)}. \end{cases} \tag{15}$$

Since there are only two absorbing sets, we have that the stochastic potential of $\overrightarrow{BB}$ is given by

$$r(\overrightarrow{AA}, \overrightarrow{BB}) = n - m^{AB}.$$

Second, consider the transition from $\overrightarrow{BB}$ to $\overrightarrow{AA}$.

Note that fewer mutations are required for B-agents with the higher constraint to switch. Thus, to obtain the minimum number of mutations required for this transition, we start with agents in the high-constraint group. Denote by m_h^{BA} the minimum number of mutations required for B-agents in N_h to switch.

Whenever $m_h^{BA} \leq n - k^h - 1$, B-agents in N_h will always stay since there are sufficient other B-agents for them to link to. Thus, we turn to the case where $m_h^{BA} > n - k^h - 1$, i.e. B-agents have to link to both A- and B-agents. Now we have to determine the payoff B-agents get when they switch to A. Thus, we have to consider two sub-cases: i) A-agents have to link to both A- and B-agents after they switch, i.e. $m_h^{BA} < k^h$, and ii) A-agents can fill all their slots with other A-agents, i.e. $m_h^{BA} \geq k^h$.

In sub-case i), the switching threshold for B-agents is given by $m_h^{BA} = \lceil M_h^1 \rceil - 1$ according to Table 1. Solving $\lceil M_h^1 \rceil - 1 < k^h$ yields that $k^h \geq \frac{(b-d)n}{a+b-2d} + \frac{a-c}{a+b-2d}$.

In sub-case ii), the switching threshold for B-agents is $m_h^{BA} = \lceil M_h^2 \rceil - 1$ according to Table 1. Then by solving $\lceil M_h^2 \rceil - 1 \geq k^h$, we obtain that $k^h < \frac{(b-d)n}{a+b-2d}$.[19]

It remains to be classified what happens in the range $k^h \in \left[\frac{(b-d)n}{a+b-2d}, \frac{(b-d)n}{a+b-2d} + \frac{a-c}{a+b-2d}\right)$. Assume that $m_h^{BA} < k^h$. Since the number of A-agents is less than the constraint, A-agents will have to link to both A- and B-agents. Attending Table 1 reveals that for B-agents to switch requires

[19]Note that k^h is a positive integer. If $k^h > \lceil M_h^1 \rceil - 1$, then $k^h \geq M_h^1$. Similarly, if $k^h \leq \lceil M_h^2 \rceil - 1$, then $k^h < M_h^2$.

that $m_h^{BA} \geq \lceil M_h^1 \rceil - 1$, which can in turn be written as $k^h \geq \frac{(b-d)n}{a+b-2d} + \frac{a-c}{a+b-2d}$. This lies out of our interval, yielding a contradiction. Thus, we consider $m_h^{BA} \geq k^h$. Now observe that for $m_h^{BA} = k^h$, we have that $v_i(A, k^h+1) \geq v_i(B, k^h)$ holds, provided that $k^h \geq \frac{(n-1)(b-d)}{a+b-2d}$. Because $k^h \in \left[\frac{(b-d)n}{a+b-2d}, \frac{(b-d)n}{a+b-2d} + \frac{a-c}{a+b-2d}\right)$, exactly k^h mutation are sufficient for the transition in this range. In summary, we have that

$$m_h^{BA} = \begin{cases} \lceil M_h^2 \rceil - 1 & \text{if } k^h < \frac{n(b-d)}{a+b-2d}. \\ k^h & \text{if } k^h \in \left[\frac{n(b-d)}{a+b-2d}, \frac{n(b-d)+(a-c)}{a+b-2d}\right). \\ \lceil M_h^1 \rceil - 1 & \text{if } k^h \geq \frac{n(b-d)+(a-c)}{a+b-2d}. \end{cases} \tag{16}$$

Now, observe that $k^h = \lceil M_h^2 \rceil - 1$ if $k^h < M_h^2 \leq k^h + 1$. This holds for $\frac{(n-1)(b-d)}{a+b-2d} \leq k^h < \frac{n(b-d)}{a+b-2d}$. Similarly, we find that $k^h = \lceil M_h^1 \rceil - 1$ if $k^h < M_h^1 \leq k^h + 1$ which can in turn be written as $k^h \in [\frac{(n-1)(b-d)}{a+b-2d}, \frac{n(b-d)}{a+b-2d} + \frac{(a-c)}{a+b-2d})$. Thus, the equation (16) is equivalent to

$$m_h^{BA} = \begin{cases} \lceil M_h^2 \rceil - 1 & \text{if } k^h < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_h^1 \rceil - 1 & \text{if } k^h \geq \frac{(n-1)(b-d)}{a+b-2d}. \end{cases} \tag{17}$$

Moreover, we find that $\lceil M_h^2 \rceil - 1 > n - k^h - 1$ holds for any k^h, and $\lceil M_h^1 \rceil - 1 > n - k^h - 1$ Whenever $k^h > \frac{(n-1)(a-c)}{a+b-2c}$. As $\frac{(n-1)(a-c)}{a+b-2c} < \frac{(n-1)(b-d)}{a+b-2d}$, equation (17) is true in the relevant range of k^h.

Now, denote by m^{BA} the minimum number of mutations for agents among both N_ℓ and N_h to switch. To maximize the impact of the mutations, assume that all mutations occur in the low-constraint group N_ℓ. Thus, after all B-agents in N_h have switched, the maximum number of A-agents is $\min\{n, m_h^{BA} + n_h\}$. It follows that the minimum number of B-agents now is $\max\{0, n_\ell - m_h^{BA}\}$. If $n_\ell - m_h^{BA} \leq 0$, i.e. if there are no B-agents, then we have reached $\overrightarrow{AA}$ and no extra mutations are required. Thus, we have $m^{BA} = m_h^{BA}$ for the relevant range of k^h.

Consider the case where there are still $m_h^{BA} + n_h$ A-agents left after the mutation and switch, i.e. $n_\ell - m_h^{BA} > 0$. Now, we have to determine whether the number of A-agents is enough for B-agents in N_ℓ to switch. Following the same argument as above, the switching threshold for B-agents in

N_ℓ is given by

$$m_\ell^{BA} = \begin{cases} \lceil M_\ell^2 \rceil - 1 & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_\ell^1 \rceil - 1 & \text{if } k^\ell \geq \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

where m_ℓ^{BA} is the minimum number of A-agents required for agents in N_ℓ to switch to A. It follows that if $m_h^{BA} + n_h > m_\ell^{BA}$, then no extra mutation are required for this transition. If $m_h^{BA} + n_h \leq m_\ell^{BA}$, an additional $m_\ell^{BA} - (m_h^{BA} + n_h)$ mutations are needed. Then total number of mutations is $m_\ell^{BA} - n_h$. In summary, the minimum number of mutations required is

$$m^{BA} = \max\{m_h^{BA}, m_\ell^{BA} - n_h\} \tag{18}$$

Since there are only two absorbing sets, we have that the stochastic potential of $\overrightarrow{BB}$ for the relevant ranges of k^ℓ and k^h is given by

$$r(\overrightarrow{BB}, \overrightarrow{AA}) = m^{BA}.$$

Having characterized the stochastic potentials of the absorbing sets, we now proceed to identify the set of stochastically stable states $\mathcal{S}^{\star\star\star}$ for the various ranges of k^ℓ and k^h. Denote by $\Delta(k^\ell, k^h)$ the difference between the stochastic potentials of $\overrightarrow{BB}$ and $\overrightarrow{AA}$, i.e. $\Delta(k^\ell, k^h) = r(\overrightarrow{AA}, \overrightarrow{BB}) - r(\overrightarrow{BB}, \overrightarrow{AA})$. If $\Delta(k^\ell, k^h) > 0$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{AA}$; if $\Delta(k^\ell, k^h) = 0$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{AA} \cup \overrightarrow{BB}$, and if $\Delta(k^\ell, k^h) < 0$, then $\mathcal{S}^{\star\star\star} = \overrightarrow{BB}$.

First, consider the case where $k^\ell < k^h < \frac{(n-1)(b-d)}{a+b-2d}$. We have obtained above that $r(\overrightarrow{BB}, \overrightarrow{AA}) = \max\{\lceil M_h^2 \rceil - 1, \lceil M_\ell^2 \rceil - 1 - n_h\}$ and $r(\overrightarrow{AA}, \overrightarrow{BB}) = n - \lfloor M_\ell^2 \rfloor$ for the relevant ranges of k^ℓ and k^h. Thus, we have that

$$\begin{aligned} \Delta(k^\ell, k^h) &= \min\{n - \lfloor M_\ell^2 \rfloor - \lceil M_h^2 \rceil + 1, n + n_h + 1 - \lfloor M_\ell^2 \rfloor - \lceil M_\ell^2 \rceil\} \\ &= \min\{\left\lceil \frac{a-d}{b-d} \cdot k^\ell \right\rceil + \left\lfloor \frac{a-d}{b-d} \cdot k^h \right\rfloor - n + 1, \left\lceil \frac{a-d}{b-d} \cdot k^\ell \right\rceil + \left\lfloor \frac{a-d}{b-d} \cdot k^\ell \right\rfloor - n_\ell + 1\}. \end{aligned}$$

One can check that $n - \lfloor M_\ell^2 \rfloor - \lfloor M_h^2 \rfloor + 1 < 0$ holds whenever $k^\ell < k^h < \frac{(n-1)(b-d)}{a+b-2d}$, which implies that $\Delta(k^\ell, k^h) < 0$. Thus, $\mathcal{S}^{\star\star\star} = \overrightarrow{BB}$. In this case, the two thresholds in the proposition are given by $\overline{\overline{k^\ell}} = \underline{\underline{k^\ell}} = \frac{(n-1)(b-d)}{a+b-2d}$.

Second, consider the case where $k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d} \leq k^h \leq \underline{k^h}$. We have that $r(\overrightarrow{BB}, \overrightarrow{AA}) = \max\{\lceil M_h^1 \rceil - 1, \lceil M_\ell^2 \rceil - 1 - n_h\}$, and $r(\overrightarrow{AA}, \overrightarrow{BB}) = n - \lfloor M_\ell^2 \rfloor$. Thus,

$$\begin{aligned}
\Delta(k^\ell, k^h) &= \min\{n - \lfloor M_\ell^2 \rfloor - \lceil M_h^1 \rceil + 1, n + n_h + 1 - \lfloor M_\ell^2 \rfloor - \lceil M_\ell^2 \rceil\} \\
&= \min\{\left\lceil \frac{a-d}{b-d} \cdot k^\ell \right\rceil + \left\lfloor \frac{k^h(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rfloor, \left\lceil \frac{a-d}{b-d} \cdot k^\ell \right\rceil + \left\lfloor \frac{a-d}{b-d} \cdot k^\ell \right\rfloor - n_\ell + 1\} \\
&:= \min\{\phi(k^\ell, k^h), \psi(k^\ell, k^h)\}.
\end{aligned}$$

Given that $b > a > c > d$, $\Delta(k^\ell, k^h)$ is weakly increasing in both k^ℓ and k^h. Thus, $\Delta(k^\ell, k^h)$ obtains its minimum at the boundary where $k^\ell = 1$ and $k^h = \frac{(n-1)(b-d)}{a+b-2d}$. At this point, we have that

$$\begin{aligned}
&\phi(1, \frac{(n-1)(b-d)}{a+b-2d}) = \left\lceil \frac{a-d}{b-d} \right\rceil - \left\lfloor \frac{(n-1)(b-d)}{a+b-2d} \right\rfloor \leq 0; \\
&\psi(1, \frac{(n-1)(b-d)}{a+b-2d}) = 2 - n_\ell \leq 0.
\end{aligned}$$

Thus, we have that $\Delta(1, \frac{(n-1)(b-d)}{a+b-2d}) \leq 0$. [20] When n is sufficiently large, we have that $\Delta(1, \frac{(n-1)(b-d)}{a+b-2d})$ is strictly negative.

We now assess the maximum of $\Delta(k^\ell, k^h)$, which is obtained at the boundary where $k^\ell = \frac{(n-1)(b-d)}{a+b-2d}$ and $k^h = \underline{k^h}$. We have that

$$\begin{aligned}
&\phi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) = \left\lceil n_\ell - \frac{(n-1)(b-d)}{a+b-2d} \right\rceil; \\
&\psi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) = \left\lceil \frac{(n-1)(a-d)}{a+b-2d} \right\rceil + \left\lfloor \frac{(n-1)(a-d)}{a+b-2d} \right\rfloor - n_\ell + 1.
\end{aligned}$$

We find that $\phi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) = 0$ if $\frac{(n-1)(b-d)}{a+b-2d} - 1 < n_\ell \leq \frac{(n-1)(b-d)}{a+b-2d}$. Thus, we have that $\phi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) < 0$ holds whenever $n_\ell \leq \frac{(n-1)(b-d)}{a+b-2d} - 1$ and $\phi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) > 0$ holds whenever $n_\ell > \frac{(n-1)(b-d)}{a+b-2d}$. Similarly, we find that $\psi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) = 0$ whenever $\frac{2(n-1)(a-d)}{a+b-2d} \leq n_\ell < \frac{2(n-1)(a-d)}{a+b-2d} + 2$. Thus, we have that $\psi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) < 0$ holds whenever $n_\ell \geq \frac{2(n-1)(b-d)}{a+b-2d} + 2$ and

[20] Notice that $\Delta(1, \frac{(n-1)(b-d)}{a+b-2d}) = 0$ hold if and only if $n = 3$ and $n_\ell = 2$, otherwise, $\Delta(1, \frac{(n-1)(b-d)}{a+b-2d}) < 0$. Furthermore, note that when $n = 3$ and $n_\ell = 2$, we have that $k_\ell = 1$ and $k_h = 2$. The transition from $\overrightarrow{AA}$ to $\overrightarrow{BB}$ requires one mutation and the transition from $\overrightarrow{BB}$ to $\overrightarrow{AA}$ requires two mutations. Thus, $\overrightarrow{BB}$ is the unique set of stochastically stable states. Thus, in the main context, we only discuss the case when n is sufficiently large.

$\psi(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) > 0$ holds whenever $n_\ell < \frac{2(n-1)(b-d)}{a+b-2d}$. In summary, we have that

$$\begin{cases} \Delta(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) < 0 & \text{if } n_\ell \in (2, \frac{(n-1)(b-d)}{a+b-2d} - 1] \cup [\frac{2(n-1)(a-d)}{a+b-2d} + 2, n-1); \\ \Delta(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) = 0 & \text{if } n_\ell \in (\frac{(n-1)(b-d)}{a+b-2d} - 1, \frac{(n-1)(b-d)}{a+b-2d}] \cup [\frac{2(n-1)(a-d)}{a+b-2d}, \frac{2(n-1)(a-d)}{a+b-2d} + 2); \\ \Delta(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) > 0 & \text{if } n_\ell \in (\frac{(n-1)(b-d)}{a+b-2d}, \frac{2(n-1)(a-d)}{a+b-2d}). \end{cases}$$

Consequently, whenever $n_\ell \in (2, \frac{(n-1)(b-d)}{a+b-2d} - 1] \cup [\frac{2(n-1)(a-d)}{a+b-2d} + 2, n-1)$, we have that $\Delta(k^\ell, k^h) < 0$ for any k^ℓ and k^h in the relevant ranges , which implies that $\mathcal{S}^{\star\star\star} = \overrightarrow{BB}$. In this case, the two thresholds in the proposition are given by $\overline{\overline{k^\ell}} = \underline{\underline{k^\ell}} = \frac{(n-1)(b-d)}{a+b-2d}$.

It follows that if $n_\ell \in (\frac{(n-1)(b-d)}{a+b-2d} - 1, \frac{(n-1)(b-d)}{a+b-2d}] \cup [\frac{2(n-1)(a-d)}{a+b-2d}, \frac{2(n-1)(a-d)}{a+b-2d} + 2)$, we have that $\Delta(k^\ell, k^h) = 0$ holds if and only if $k^\ell = \frac{(n-1)(b-d)}{a+b-2d}$ and $k^h = \underline{k^h}$, which implies $\mathcal{S}^{\star\star\star} = \overrightarrow{BB} \cup \overrightarrow{AA}$. Furthermore, we have that $\Delta(k^\ell, k^h) < 0$ for any pair of k^ℓ and k^h such that $k^\ell < \frac{(n-1)(b-d)}{a+b-2d} \leq k^h < \underline{k^h}$, which implies that $\mathcal{S}^{\star\star\star} = \overrightarrow{BB}$. The two thresholds in this case are thus given by $\overline{\overline{k^\ell}} = \underline{\underline{k^\ell}} = \frac{(n-1)(b-d)}{a+b-2d}$.

Moreover, if $n_\ell \in (\frac{(n-1)(b-d)}{a+b-2d}, \frac{2(n-1)(a-d)}{a+b-2d})$, we have that the maximum of $\Delta(k^\ell, k^h)$ is positive, i.e. $\Delta(\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}) > 0$ and the minimum is negative, i.e. $\Delta(1, \frac{(n-1)(b-d)}{a+b-2d}) < 0$. Thus, for each $k^h \in [\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h})$, there exists a corresponding interval of k^ℓ, such that for any k^ℓ in this interval we have that $\Delta(k^\ell, k^h) = 0$. Note that $\Delta(k^\ell, k^h)$ is weakly increasing in both k^ℓ and k^h. We have that $\Delta(k^\ell, k^h) < 0$ if k^ℓ falls below this interval and $\Delta(k^\ell, k^h) > 0$ if k^ℓ falls above. Therefore, for each $k^h \in [\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h})$, we have that

$$\mathcal{S}^{\star\star\star} = \begin{cases} \overrightarrow{BB}, & \text{if } k^\ell < \underline{\underline{k^\ell}}. \\ \overrightarrow{BB} \cup \overrightarrow{AA}, & \text{if } k^\ell \in [\underline{\underline{k^\ell}}, \overline{\overline{k^\ell}}]. \\ \overrightarrow{AA}, & \text{if } k^\ell > \overline{\overline{k^\ell}}. \end{cases}$$

where $\overline{\overline{k^\ell}}$ and $\underline{\underline{k^\ell}}$ are the two thresholds which are given by the upper and lower boundaries of this interval respectively.

Now, consider the case where $\frac{(n-1)(b-d)}{a+b-2d} < k^\ell < k^h \leq \underline{k^h}$. we have that $r(\overrightarrow{BB}, \overrightarrow{AA}) =$

$\max\{\lceil M_h^1 \rceil - 1, \lceil M_\ell^1 \rceil - 1 - n_h\}$, and $r(\overrightarrow{AA}, \overrightarrow{BB}) = n - \lfloor M_\ell^1 \rfloor$. Thus,

$$\begin{aligned}
\Delta(k^\ell, k^h) &= \min\{n - \lfloor M_\ell^1 \rfloor - \lceil M_h^1 \rceil + 1, n + n_h + 1 - \lfloor M_\ell^1 \rfloor - \lceil M_\ell^1 \rceil\} \\
&= \min\{n + \left\lceil \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rceil + \left\lfloor \frac{k^h(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rfloor - 1, \\
&\quad n + n_h + \left\lceil \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rceil + \left\lfloor \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rfloor - 1\} \\
&:= \min\{\phi(k^\ell, k^h), \psi(k^\ell, k^h)\}.
\end{aligned}$$

As above, $\Delta(k^\ell, k^h)$ is weakly increasing in both k^ℓ and k^h. Thus, $\Delta(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) < \Delta(k^\ell, k^h) < \Delta(\underline{k^h}, \underline{k^h})$. One can check that $\Delta(\underline{k^h}, \underline{k^h}) > 0$ hold since both $\phi(\underline{k^h}, \underline{k^h})$ and $\psi(\underline{k^h}, \underline{k^h})$ are strictly positive. We now assess the sign of $\Delta(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d})$. Note that $\Delta(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) = \min\{\phi(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}), \psi(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d})\}$ where

$$\phi(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) = n - \left\lceil \frac{(n-1)(b-d)}{a+b-2d} \right\rceil - \left\lfloor \frac{(n-1)(b-d)}{a+b-2d} \right\rfloor - 1 < 0.$$

and

$$\psi(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) = n + n_h - \left\lceil \frac{(n-1)(b-d)}{a+b-2d} \right\rceil - \left\lfloor \frac{(n-1)(b-d)}{a+b-2d} \right\rfloor - 1.$$

Since $\phi(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) < 0$ holds, we have that $\Delta(\frac{(n-1)(b-d)}{a+b-2d}, \frac{(n-1)(b-d)}{a+b-2d}) < 0$. Thus, for each $k^h \in (\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}]$, there exists a corresponding interval of k^ℓ, such that for any k^ℓ in this interval we have $\Delta(k^\ell, k^h) = 0$. As above, We have that $\Delta(k^\ell, k^h) < 0$ if k^ℓ falls below this interval and $\Delta(k^\ell, k^h) > 0$ if k^ℓ falls above. Therefore, for each $k^h \in (\frac{(n-1)(b-d)}{a+b-2d}, \underline{k^h}]$, we have that

$$\mathcal{S}^{\star\star\star} = \begin{cases} \overrightarrow{BB}, & \text{if } k^\ell < \underline{\underline{k^\ell}}. \\ \overrightarrow{BB} \cup \overrightarrow{AA}, & \text{if } k^\ell \in [\underline{\underline{k^\ell}}, \overline{\overline{k^\ell}}]. \\ \overrightarrow{AA}, & \text{if } k^\ell > \overline{\overline{k^\ell}}. \end{cases}$$

where $\overline{\overline{k^\ell}}$ and $\underline{\underline{k^\ell}}$ are the two thresholds which are given by the upper and lower boundaries of this interval respectively.

Finally, consider the case where $k^\ell \geq \overline{k^\ell}$ and $k^h > \underline{k^h}$. In this case, we also have that $r(\overrightarrow{BB}, \overrightarrow{AA}) = \max\{\lceil M_h^1 \rceil - 1, \lceil M_\ell^1 \rceil - 1 - n_h\}$, and $r(\overrightarrow{AA}, \overrightarrow{BB}) = n - \lfloor M_\ell^1 \rfloor$. Thus,

$$
\begin{aligned}
\Delta(k^\ell, k^h) &= \min\{n - \lfloor M_\ell^1 \rfloor - \lceil M_h^1 \rceil + 1, n + n_h + 1 - \lfloor M_\ell^1 \rfloor - \lceil M_\ell^1 \rceil\} \\
&= \min\{n + \left\lceil \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rceil + \left\lfloor \frac{k^h(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rfloor - 1, \\
&\quad n + n_h + \left\lceil \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rceil + \left\lfloor \frac{k^\ell(c-d) - (n-1)(b-d)}{a+b-c-d} \right\rfloor - 1\} \\
&:= \min\{\phi(k^\ell, k^h), \psi(k^\ell, k^h)\}.
\end{aligned}
$$

One can check that both $\phi(k^\ell, k^h)$ and $\psi(k^\ell, k^h)$ are strictly positive if $k^\ell \geq \overline{k^\ell}$ and $k^h > \underline{k^h}$. Therefore, we have that $\Delta(k^\ell, k^h) > 0$ and consequently, $\mathcal{S}^{\star\star\star} = \overrightarrow{AA}$ for any k^ℓ and k^h with $k^\ell \geq \overline{k^\ell}$ and $k^h > \underline{k^h}$. In this case, the two thresholds in the proposition are given by $\overline{\overline{k^\ell}} = \underline{\underline{k^\ell}} = \overline{k^\ell}$.
□

Proof of Proposition 5.3. First, note that according to Proposition 5.1, for any k^ℓ and k^h with $k^\ell < \overline{k^\ell}$ and $k^h > \underline{k^h}$ there are three absorbing sets $\overrightarrow{AA}$, $\overrightarrow{BB}$ and $\overrightarrow{BA}$. The proof proceeds by using techniques by Young (1993) and Kandori et al. (1993), which include three steps: i) calculate the resistance of transition from one absorbing set to another; ii) calculate the stochastic potential of each absorbing set, and iii) compare the stochastic potentials and find the smallest one.

i) Calculate the resistances of transitions.

First, we consider the transition from $\overrightarrow{AA}$ to $\overrightarrow{BA}$. Following the same argument as in the proof of Proposition 5.2, the minimum number of mutations for the transitions of agents in N_ℓ is given by

$$
n - m_\ell^{AB} = \begin{cases} n - \lfloor M_\ell^2 \rfloor, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^1 \rfloor, & \text{if } \frac{(n-1)(b-d)}{a+b-2d} < k^\ell < \overline{k^\ell}. \end{cases}
$$

Assume that all mutations occur in N_ℓ. After the mutations and consequent switches have occured, we have reached a state in $\overrightarrow{BA}$. Note that $\overrightarrow{BA}$ is absorbing. Thus, no agent will switch without further mutations. Therefore, the resistance of the transition from $\overrightarrow{AA}$ to $\overrightarrow{BA}$ is given by

$$r(\overrightarrow{AA}, \overrightarrow{BA}) = \begin{cases} n - \lfloor M_\ell^2 \rfloor, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^1 \rfloor, & \text{if } \frac{(n-1)(b-d)}{a+b-2d} < k^\ell < \overline{k^\ell}. \end{cases} \tag{19}$$

Second, we consider the transition from $\overrightarrow{BA}$ to $\overrightarrow{AA}$. Following the same argument as in the proof of Proposition 5.2, we have that the minimum number of A-agents required for agents in N_ℓ to switch from B to A is given by

$$m_\ell^{BA} = \begin{cases} \lceil M_\ell^2 \rceil - 1, & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_\ell^1 \rceil - 1, & \text{if } \frac{(n-1)(b-d)}{a+b-2d} \leq k^\ell < \overline{k^\ell}. \end{cases}$$

Since there are already n_h A-agents before the mutations, the number of mutations required for agents in N_ℓ to switch is $m_\ell^{BA} - n_h$. Thus, for any $k^h > \underline{k^h}$, the resistance of the transition from $\overrightarrow{BA}$ to $\overrightarrow{AA}$ is given by

$$r(\overrightarrow{BA}, \overrightarrow{AA}) = \begin{cases} \lceil M_\ell^2 \rceil - n_h - 1, & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_\ell^1 \rceil - n_h - 1, & \text{if } \frac{(n-1)(b-d)}{a+b-2d} \leq k^\ell < \overline{k^\ell}. \end{cases} \tag{20}$$

Third, consider the transition from $\overrightarrow{BB}$ to $\overrightarrow{BA}$. Consider agents in N_h. Attending to Table 1 reveals that the minimum number of mutations required for agents in N_h to switch from B to A is given by $m_h^{BA} = \lceil M_h^1 \rceil - 1$. Assume that all mutations occur in N_h. After the mutations and consequent switches have occured, we have reached a state in $\overrightarrow{BA}$. Since $\overrightarrow{BA}$ is absorbing, no agent will switch without further mutations. Thus, the resistance of the transition from $\overrightarrow{BB}$ to $\overrightarrow{BA}$ is given by

$$r(\overrightarrow{BB}, \overrightarrow{BA}) = \lceil M_h^1 \rceil - 1. \tag{21}$$

Next, consider the transition from $\overrightarrow{BA}$ to $\overrightarrow{BB}$. Denote by $n - m_h^{AB}$ the minimum number of B-agents required for agents in N_h to switch form action A to B. Thus, m_h^{AB} is the maximum number of A-agents allowed for this transition. According to Table 1, we have that $m_h^{AB} = \lfloor M_h^1 \rfloor$. Since there are already n_ℓ B-agents before the mutations, the minimum number of mutations required is thus $n - m_h^{AB} - n_\ell$. Hence, the resistance of the transition from $\overrightarrow{BA}$ to $\overrightarrow{BB}$ is given by

$$r(\overrightarrow{BA}, \overrightarrow{BB}) = n_h - \lfloor M_h^1 \rfloor . \tag{22}$$

Now, consider the transition from $\overrightarrow{AA}$ to $\overrightarrow{BB}$. Following the same argument as in the proof of Proposition 5.2, the minimum number of mutations for the transitions of agents in N_ℓ is given by

$$n - m_\ell^{AB} = \begin{cases} n - \lfloor M_\ell^2 \rfloor , & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^1 \rfloor , & \text{if } \frac{(n-1)(b-d)}{a+b-2d} < k^\ell < \overline{k^\ell}. \end{cases}$$

We now assess the largest number of B-agents after the mutations and switches (excluding switches among agents in N_h for now), i.e. agents who have mutated and agents in N_ℓ who have switched to B. For this, assume all mutations occur in N_h. Thus, the largest number of B-agents is $n_\ell + n - m_\ell^{AB}$. It follows that the minimum number of remaining A-agents in N_h is $m_\ell^{AB} - n_\ell$. Consider the transitions of agents in N_h now. As above, the maximum number of A-agents allowed for the transitions of agents in N_h is $m_h^{AB} = \lfloor M_h^1 \rfloor$ whenever $k^h > \underline{k^h}$. If the number of A-agents $m_\ell^{AB} - n_\ell$ is less than m_h^{AB}, then agents in N_h will switch without further mutations. In this case, the resistance of the transition is given by

$$r(\overrightarrow{AA}, \overrightarrow{BB}) = \begin{cases} n - \lfloor M_\ell^2 \rfloor , & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^1 \rfloor , & \text{if } \frac{(n-1)(b-d)}{a+b-2d} < k^\ell < \overline{k^\ell}. \end{cases}$$

Otherwise, if $m_\ell^{AB} - n_\ell \geq m_h^{AB}$, additional mutations are needed. In this case, assume that the minimum number of mutations required is x. Then, we have that $n - (n_\ell + x) < m_h^{AB}$ must hold, i.e. the number of A-agents left after the mutations and switched in N_ℓ is less than the switching threshold. Thus, we have that $x > n_h - m_h^{AB}$. Since x is the minimum number, $x = n_h - m_h^{AB} + 1$. One can check that $n_h - m_h^{AB} + 1 > n - m_\ell^{AB}$. Therefore, we have that

$$r(\overrightarrow{AA}, \overrightarrow{BB}) \geq n - m_\ell^{AB} = r(\overrightarrow{AA}, \overrightarrow{BA}). \tag{23}$$

Finally, consider the transition from $\overrightarrow{BB}$ to $\overrightarrow{AA}$. As above, the minimum number of mutations required for agents in N_h to switch from B to A is given by $m_h^{BA} = \lceil M_h^1 \rceil - 1$.

To maximize the impact of the mutations, assume that all mutations occur in the low-constraint group N_ℓ. Thus, after all B-agents in N_h have switched, the maximum number of A-agents is $\min\{n, m_h^{BA} + n_h\}$. It follows that the minimum number of B-agents now is $\max\{0, n_\ell - m_h^{BA}\}$. If $n_\ell - m_h^{BA} \leq 0$, i.e. if there are no B-agents, then we have reached $\overrightarrow{AA}$ and no extra mutations are required. Thus, we have $m^{BA} = m_h^{BA}$ for the relevant range of k^h.

Consider the case where there are still $m_h^{BA} + n_h$ A-agents left after the mutation and switch, i.e. $n_\ell - m_h^{BA} > 0$. Now, we have to determine whether the number of A-agents is enough for B-agents in N_ℓ to switch. Following the same argument as above, the switching threshold for B-agents in N_ℓ is given by

$$m_\ell^{BA} = \begin{cases} \lceil M_\ell^2 \rceil - 1 & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_\ell^1 \rceil - 1 & \text{if } k^\ell \geq \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

where m_ℓ^{BA} is the minimum number of A-agents required for agents in N_ℓ to switch to A. It follows that if $m_h^{BA} + n_h > m_\ell^{BA}$, then no extra mutation are required for this transition. In this case, we have that

$$r(\overrightarrow{BB}, \overrightarrow{AA}) = \lceil M_h^1 \rceil - 1.$$

If $m_h^{BA} + n_h \leq m_\ell^{BA}$, an additional $m_\ell^{BA} - (m_h^{BA} + n_h)$ mutations are needed. Then total number of mutations is $m_\ell^{BA} - n_h$. In this case, we have that

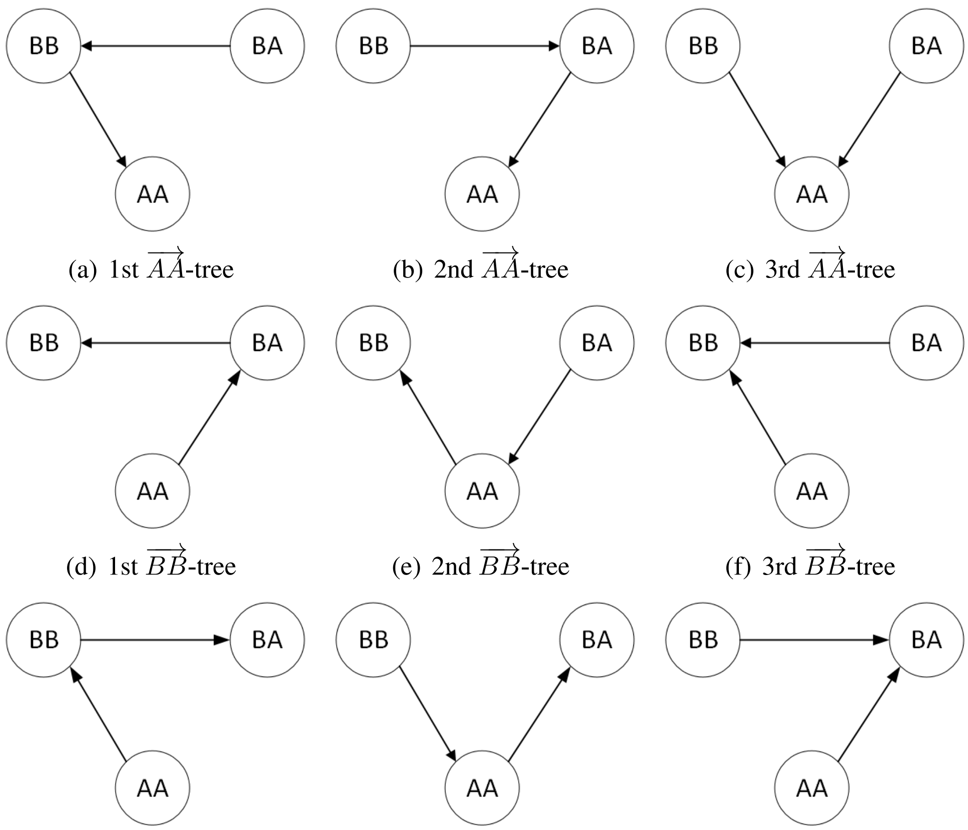
$$r(\overrightarrow{BB}, \overrightarrow{AA}) = \begin{cases} \lceil M_\ell^2 \rceil - n_h - 1 & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \\ \lceil M_\ell^1 \rceil - n_h - 1 & \text{if } k^\ell \geq \frac{(n-1)(b-d)}{a+b-2d}. \end{cases}$$

One can check that in both cases, we have that

$$r(\overrightarrow{BB}, \overrightarrow{AA}) \geq \lceil M_h^1 \rceil - 1 = r(\overrightarrow{BB}, \overrightarrow{BA}). \tag{24}$$

ii) Calculate the stochastic potential of each absorbing set.

Having obtained the resistances of transitions, we are now able to compute the stochastic potentials of each absorbing set. We denote by $r_j(S_i^{\star\star})$ the resistance of the j-th $S_i^{\star\star}$-tree. Figure 5 depicts all possible $\overrightarrow{AA}$, $\overrightarrow{BB}$ and $\overrightarrow{BA}$-trees.



(a) 1st $\overrightarrow{AA}$-tree (b) 2nd $\overrightarrow{AA}$-tree (c) 3rd $\overrightarrow{AA}$-tree

(d) 1st $\overrightarrow{BB}$-tree (e) 2nd $\overrightarrow{BB}$-tree (f) 3rd $\overrightarrow{BB}$-tree

(g) 1st $\overrightarrow{BA}$-tree (h) 2nd $\overrightarrow{BA}$-tree (i) 3rd $\overrightarrow{BA}$-tree

Figure 5: All $S_i^{\star\star}$-trees

First, consider all $\overrightarrow{AA}$-trees depicted as sub-figures 5(a) to 5(c) in Figure 5. The resistances of these trees are given by

$$r_1(\overrightarrow{AA}) = r(\overrightarrow{BA}, \overrightarrow{BB}) + r(\overrightarrow{BB}, \overrightarrow{AA}),$$
$$r_2(\overrightarrow{AA}) = r(\overrightarrow{BB}, \overrightarrow{BA}) + r(\overrightarrow{BA}, \overrightarrow{AA}),$$
$$r_3(\overrightarrow{AA}) = r(\overrightarrow{BB}, \overrightarrow{AA}) + r(\overrightarrow{BA}, \overrightarrow{AA}).$$

Given inequality (24), we have that $r_3(\overrightarrow{AA}) \geq r_2(\overrightarrow{AA})$. Thus, the stochastic potential of $\overrightarrow{AA}$ is given by

$$\gamma(\overrightarrow{AA}) = \min\{r_1(\overrightarrow{AA}), r_2(\overrightarrow{AA})\}. \tag{25}$$

Now, consider all $\overrightarrow{BB}$-trees depicted as sub-figures 5(d) to 5(f) in Figure 5. The resistances of these trees are given by

$$\begin{aligned}
r_1(\overrightarrow{BB}) &= r(\overrightarrow{AA}, \overrightarrow{BA}) + r(\overrightarrow{BA}, \overrightarrow{BB}), \\
r_2(\overrightarrow{BB}) &= r(\overrightarrow{BA}, \overrightarrow{AA}) + r(\overrightarrow{AA}, \overrightarrow{BB}), \\
r_3(\overrightarrow{BB}) &= r(\overrightarrow{BA}, \overrightarrow{BB}) + r(\overrightarrow{AA}, \overrightarrow{BB}).
\end{aligned}$$

Given inequality (23), we have that $r_3(\overrightarrow{BB}) \geq r_1(\overrightarrow{BB})$. Thus, the stochastic potential of $\overrightarrow{BB}$ is given by

$$\gamma(\overrightarrow{BB}) = \min\{r_1(\overrightarrow{BB}), r_2(\overrightarrow{BB})\}. \tag{26}$$

Finally, consider all $\overrightarrow{BB}$-trees depicted as sub-figures 5(g) to 5(i) in Figure 5. The resistances of these trees are given by

$$\begin{aligned}
r_1(\overrightarrow{BA}) &= r(\overrightarrow{AA}, \overrightarrow{BB}) + r(\overrightarrow{BB}, \overrightarrow{BA}), \\
r_2(\overrightarrow{BA}) &= r(\overrightarrow{BB}, \overrightarrow{AA}) + r(\overrightarrow{AA}, \overrightarrow{BA}), \\
r_3(\overrightarrow{BA}) &= r(\overrightarrow{BB}, \overrightarrow{BA}) + r(\overrightarrow{AA}, \overrightarrow{BA}).
\end{aligned}$$

Given the two inequalities (23) and (24) , we have that $r_1(\overrightarrow{BA}) \geq r_3(\overrightarrow{BA})$ and $r_2(\overrightarrow{BA}) \geq r_3(\overrightarrow{BA})$. Thus, the stochastic potential of $\overrightarrow{BA}$ is given by

$$\gamma(\overrightarrow{BA}) = r_3(\overrightarrow{BA}) = \begin{cases} n - \lfloor M_\ell^1 \rfloor + \lceil M_h^1 \rceil - 1 & \text{if } k^\ell \geq \frac{(n-1)(b-d)}{a+b-2d}. \\ n - \lfloor M_\ell^2 \rfloor + \lceil M_h^1 \rceil - 1 & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \end{cases} \tag{27}$$

iii) Find the region of k^ℓ and k^h such that the stochastic potential of $\overrightarrow{BA}$ is the smallest.

Having obtained the stochastic potentials, we now move on to find the regions of k^ℓ and k^h where the stochastic potential of $\overrightarrow{BA}$ is the smallest. To do so, both $\gamma(\overrightarrow{BA}) \leq \gamma(\overrightarrow{AA})$ and $\gamma(\overrightarrow{BA}) \leq \gamma(\overrightarrow{BB})$ must hold. Given equations (25), (26) and (27), we thus have that

$$\gamma(\overrightarrow{AA}) - \gamma(\overrightarrow{BA}) = \min\{r_1(\overrightarrow{AA}) - r_3(\overrightarrow{BA}), r_2(\overrightarrow{AA}) - r_3(\overrightarrow{BA})\} \geq 0. \tag{28}$$

and

$$\gamma(\overrightarrow{BB}) - \gamma(\overrightarrow{BA}) = \min\{r_1(\overrightarrow{BB}) - r_3(\overrightarrow{BA}), r_2(\overrightarrow{BB}) - r_3(\overrightarrow{BA})\} \geq 0. \tag{29}$$

Note that above two inequalities (28) and (29) hold if and only if the following four inequalities hold

$$\begin{aligned} r_1(\overrightarrow{AA}) \geq r_3(\overrightarrow{BA}), &\quad r_2(\overrightarrow{AA}) \geq r_3(\overrightarrow{BA}); \\ r_1(\overrightarrow{BB}) \geq r_3(\overrightarrow{BA}), &\quad r_2(\overrightarrow{BB}) \geq r_3(\overrightarrow{BA}). \end{aligned} \tag{30}$$

By substituting above equations of the resistances, we rewrite inequalities in (30) as following

$$r(\overrightarrow{BA}, \overrightarrow{BB}) + r(\overrightarrow{BB}, \overrightarrow{AA}) \geq r(\overrightarrow{BB}, \overrightarrow{BA}) + r(\overrightarrow{AA}, \overrightarrow{BA}), \tag{31a}$$

$$r(\overrightarrow{BA}, \overrightarrow{AA}) - r(\overrightarrow{AA}, \overrightarrow{BA}) \geq 0, \tag{31b}$$

$$r(\overrightarrow{BA}, \overrightarrow{BB}) - r(\overrightarrow{BB}, \overrightarrow{BA}) \geq 0, \tag{31c}$$

$$r(\overrightarrow{BA}, \overrightarrow{AA}) + r(\overrightarrow{AA}, \overrightarrow{BB}) \geq r(\overrightarrow{BB}, \overrightarrow{BA}) + r(\overrightarrow{AA}, \overrightarrow{BA}). \tag{31d}$$

Now, we substitute our results of resistances of transitions in the above four inequalities (31a) to (31d), we have that

$$n_h + 1 - \lfloor M_h^1 \rfloor - \lceil M_h^1 \rceil \geq 0, \tag{32a}$$

$$\lfloor M_\ell \rfloor + \lceil M_\ell \rceil - n - n_h - 1 \geq 0., \tag{32b}$$

$$\lfloor M_\ell \rfloor - \lfloor M_h^1 \rfloor - n_\ell \geq 0, \tag{32c}$$

$$\lceil M_\ell \rceil - \lceil M_h^1 \rceil - n_h \geq 0. \tag{32d}$$

where

$$M_\ell = \begin{cases} M_\ell^2, & \text{if } k^\ell \leq \frac{(n-1)(b-d)}{a+b-2d}. \\ M_\ell^1, & \text{if } \frac{(n-1)(b-d)}{a+b-2d} < k^\ell < \overline{k^\ell}. \end{cases}$$

Let $\Phi(k^h) = n_h + 1 - \lfloor M_h^1 \rfloor - \lceil M_h^1 \rceil$ and $\Psi(k^\ell) = \lfloor M_\ell \rfloor + \lceil M_\ell \rceil - n - n_h - 1$. We have that

$$\Phi(k^h) = n_h - 1 - \left\lfloor \frac{(n-1)(b-d) - k^h(c-d)}{a+b-c-d} \right\rfloor - \left\lceil \frac{(n-1)(b-d) - k^h(c-d)}{a+b-c-d} \right\rceil. \tag{33}$$

and

$$\Psi(k^\ell) = \begin{cases} \left\lfloor \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d} \right\rfloor + \left\lceil \frac{(n-1)(b-d)-k^\ell(c-d)}{a+b-c-d} \right\rceil - n - n_h + 1 & \text{if } k^\ell \geq \frac{(n-1)(b-d)}{a+b-2d}. \\ \left\lfloor n - \frac{a-d}{b-d} \cdot k^\ell \right\rfloor + \left\lceil n - \frac{a-d}{b-d} \cdot k^\ell \right\rceil - n - n_h - 1 & \text{if } k^\ell < \frac{(n-1)(b-d)}{a+b-2d}. \end{cases} \tag{34}$$

We find that $\Phi(k^h) = 0$ whenever $k^h \in [\frac{2(n-1)(b-d)-n_h(a+b-c-d)}{2(c-d)} - \frac{a+b-c-d}{c-d}, \frac{2(n-1)(b-d)-n_h(a+b-c-d)}{2(c-d)})$. Let $k^{h\star} \equiv \frac{2(n-1)(b-d)-n_h(a+b-c-d)}{2(c-d)} - \frac{a+b-c-d}{c-d}$. Note that $\Phi(k^h)$ is weakly increasing in k^h. Thus, for any $k^h \geq k^{h\star}$, we have that $\Phi(k^h) \geq 0$.

Moreover, $\Psi(k^\ell)$ is weakly decreasing in k^ℓ with $k^\ell < \overline{k^\ell}$. Thus, we have that

$$\Psi_{\min}(k^\ell) > \Phi(\overline{k^\ell}) = 1 - n_\ell < 0. \tag{35}$$

and

$$\Psi_{\max}(k^\ell) = \Phi(1) = n_\ell - 2 \geq 0. \tag{36}$$

Thus, there exists an interval of k^ℓ, such that for any k^ℓ in this interval we have $\Psi(k^\ell) = 0$. Since $\Psi(k^\ell)$ is weakly decreasing in k^ℓ, we have that $\Psi(k^\ell) > 0$ if k^ℓ falls below this interval and

$\Psi(k^\ell) < 0$ if k^ℓ falls above. Let $k^{\ell\star}$ equal to the upper bound of this interval. Thus, for any $k^\ell \leq k^{\ell\star}$ we have that $\Psi(k^\ell) \geq 0$.

Whenever $k^\ell \leq k^{\ell\star}$ and $k^h \geq k^{h\star}$, one can check that $\lfloor M_\ell \rfloor - \lfloor M_h^1 \rfloor - n_\ell \geq 0$ and $\lceil M_\ell \rceil - \lceil M_h^1 \rceil - n_h \geq 0$ hold. Thus, for any k^ℓ and k^h with $k^\ell \leq k^{\ell\star}$ and $k^h \geq k^{h\star}$, we have that $\gamma(\overrightarrow{BA}) \leq \gamma(\overrightarrow{AA})$ and $\gamma(\overrightarrow{BA}) \leq \gamma(\overrightarrow{BB})$. Consequently, $\overrightarrow{BA} \subseteq \mathcal{S}^{\star\star\star}$.

Now, we proceed to identify thresholds for k^ℓ and k^h such that $\overrightarrow{BA}$ is the unique set of stochastically stable states. The proof is almost the same as the argument above. The only difference is in inequalities (32a) to (32d), which now are required to be strictly positive. Then, instead of solving $\Phi(k^h) \geq 0$ and $\Psi(k^\ell) \geq 0$ as above, we now solve $\Phi(k^h) > 0$ and $\Psi(k^\ell) > 0$ for the ranges of k^ℓ and k^h. In this case, $k^{h\star\star}$ is the upper bound of the solution such that $\Phi(k^h) = 0$. Similarly, $k^{\ell\star\star}$ is now the lower bound of the solution such that $\Psi(k^\ell) \geq 0$. Consequently, for any k^ℓ and k^h with $k^\ell < k^{\ell\star\star}$ and $k^h > k^{h\star\star}$, we have that $\gamma(\overrightarrow{BA}) < \gamma(\overrightarrow{AA})$ and $\gamma(\overrightarrow{BA}) < \gamma(\overrightarrow{BB})$. Thus, $\mathcal{S}^{\star\star\star} = \overrightarrow{BA}$.
□