Norms and the evolution of leaders' followership*

Antonio Cabrales†, Esther Hauk‡

February 2022

Abstract

In this paper we model the interaction between leaders, their followers and crowd followers in a coordination game with local interaction. The steady states of a dynamic best-response process can feature a coexistence of Pareto dominant and risk dominant actions in the population. The existence of leaders and their followers, plus the local interaction, which leads to clustering, is crucial for the survival of the Pareto dominant actions. The evolution of leader and crowd followership shows that leader followership can also be locally stable around Pareto dominant leaders.

*Esther Hauk acknowledges financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centers of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S)

†Department of Economics, Universidad Carlos III de Madrid; e-mail: antonio.cabrales@uc3m. es

‡Instituto de Anàliisis Econòmico (IAE-CSIC), Move and Barcelona School of Economics, Campus UAB, 08193 Bellaterra (Barcelona), email: esther.hauk@iae.csic.es

1 Introduction

Many human activities are characterized by a coordination problem: there are several optimal actions for an individual but whether the chosen action is indeed optimal depends on what the people she interacts with are doing. Hence, multiple stable social situations can arise. Examples abound, from technological standards (Farrell and Saloner 1988, Venkatraman and Lee 2004), to the multiple equilibria in repeated games that are pervasive in human interaction (think specifically of contributions to climate change mitigation, see e.g. Lempert et al. 2006).

In this paper we examine the impact of the behavior of leaders and their followers, on the adoption of an action that is good for the collectivity. Previous literature has emphasized social conventions as a common way to deal with the issue of coordination (Young 1993, 1998, Burke and Young 2011). But how do those conventions arise? Humans, like most primates, are a species where groups are rather hierarchical, something that has important implications in our psychology (Cummins 2005), social organization (Manner and Case 2016), and health (Gilbert 2001). Some individuals take an action mostly because that is what their leaders do. Alternatively, they can take an action because they prefer to act like their peers. Bicchieri and Chavez (2010) or Krupka and Weber (2013), for example, have measured the impact of others' expectations about the "correct action" on our own choices. Our main aim is precisely to analyze the interactions of those two ways of reaching a convention. We do this in an environment where two possible conventions may arise, which differ in their societal level of welfare.

We model a game in which there are N players located in a circle. Each of them interacts with their neighborhood of size k who constitute the k closest players located evenly to her left and right. The interaction consists in playing simultaneous two person coordination game with all people in the neighborhood using always the same action which is either payoff dominant or a risk dominant. There are three types of players: Leaders (L), Leader-Followers (LF) and Crowd-Followers (CF). Leaders always take the same action independently of all other players' choices. Additionally to the material payoff

from the coordination games Leader-Followers receive a payoff of α_L when following their leader, who is the leader closest to them. The level of α_L reflects the leader's charisma. Finally, Crowd-Followers (CF) care about choosing the same action as their neighbors. Therefore, in addition to the material payoffs from the coordination games, they receive an extra amount α_C multiplied by the fraction of k closest neighbors who choose the same action as themselves, where α_C captures the weight given to this peer influence.

We analyze the game as a dynamic process in which agents adjust their actions over time. We first analyze the steady states of population choices when individuals start from an arbitrary action and best-respond to the population choices in the previous period. Clearly, the steady states are equilibria of the game. Then, when play reaches a steady state, we allow all agents except the Leaders to switch not just their actions, but also their types.

The aim of the model is to capture situations in which there are different norms competing for societal dominance. A *good* but *weak* norm (the Pareto dominant, but risk dominated one) and a *bad* but *strong* one (the Pareto dominated, but risk dominant one). There are numerous possible applications: using a clean or a polluting energy (Ang et al. 2020), the adoption of a farming technology (Müller et al. 2018), the choice of a software platform for developers (Fang et al. 2021), language adoption (Iriberri and Uriarte 2012), the spread of academic ideas (Sunstein 2000), expression of opinions on controversial social topics (Buskens et al. 2008) among others. In many of these applications Leaders and their Followers, or Crowd-Followers are crucial to which option survives, and how.

We characterize the steady states of the system for fixed types, as well as with evolving types. An important insight is that clustering is a crucial factor for the simultaneous survival of multiple norms by allowing the "good but weak" Pareto efficient norms to survive. Interestingly, the Leader-Follower players are particularly important to achieve this clustering. But even if they are useful for the survival, it is hard for them to take over.

Another insight delivered by the model is the stark asymmetry in importance between the Leader of the *good* (payoff-dominant) and *bad* (risk-dominant) norms. The risk dominant norm is guaranteed to survive in the

population so long as there is one Leader subscribed to the risk dominant action even if nobody gains from following her. However, if a Leader choosing the payoff dominant action is not sufficiently charismatic (low values of α_L) the payoff dominant norm may disappear. This norm will only take over in a cluster in part of the population if α_L is sufficiently high, and there are no risk dominant Leaders located inside the cluster.

We explore the possibility of a policy by which a social planner can remove a Leader to improve welfare. One can think of this as the planner targeting for "behavior change" by removing an influencer, in a context where she cannot change them all. This policy can only work if payoff dominant influencers are sufficiently charismatic. Since welfare is improved if the payoff dominant norm spreads further, it is clear that the target can only be a risk dominant leader. But which one should be chosen? This depends on the moment of the evolution of society this removal takes place. If a risk dominant leader is removed at the very beginning of the game, the removal can only be effective in enhancing payoff dominant play if the targeted risk dominant Leader was located between two payoff dominant Leaders. But here, the planner faces a trade-off. If these two payoff dominant Leaders are very far away, the new payoff dominant cluster would be very large if it was created. However, the chances that it is indeed created, are lowest in that case.

Suppose now the removal of a risk dominant Leader happens after a steady state with fixed types is reached, or after a steady state with evolving types is reached. Then, it is not necessarily optimal to remove a risk dominant Leader located between two payoff dominant Leaders. Sometimes it can be better to remove a risk dominant Leader who only has one payoff dominant Leader as a neighbor. This is true when payoff dominant Leader gains a very large sphere of influence due to this removal.

A general takeaway of this model is that "good but weak" (i.e. payoff dominant, but not risk dominant) social norms need clustered groups of supporters, and very charismatic leaders. These two features play an important role in explaining, for example, the survival of Apple at a time when it seemed it could have been taken over by Microsoft. Both the community of graphics designers (Holmberg, Logander, and Lindqvist 2005) and having an appealing

leader like Steve Jobs were very important in this story (Ruijuan, Wang and Hao 2020). Staying true to its aspiration of being different and its focus on innovation and design has allowed Apple to establish a credible brand identity. Steve Jobs' charisma and visions did not only result in Apple having highly dedicated employees despite Job's demanding managing style (Levy, 2000), his visionary leadership also led to a consumers' cult (Belk and Tumbat, 2005) where many Apple customers worshiped their Leader.

The remainder of the paper is organized as follows: the next section discusses the related literature. The model is laid out in Section 3 and the steady states with fixed types and evolving types are analyzed in Sections 4 and 5 respectively. Section 6 discusses the importance of Leaders for the survival of risk versus payoff dominant norms and answers the question which Leader should be removed to enhance payoff dominant play. Section 7 concludes and suggests directions for future research.

2 Related literature

There are several important strands of the literature that connect to our work. Most obviously, Acemoglu and Jackson (2015, 2017) have explored the role of social norms and leadership in coordination games. They follow on the seminal contributions of Young (1993, 1998), Binmore and Samuelson (1994).[1] Our contribution to that literature is twofold. On the one hand we emphasize the local aspect of social norms enforcement, and the possibility of multiple social norms in steady state through local clustering. On the other hand, we emphasize the importance of agents who follow leaders, versus those who simply follow crowds, and their interaction.

Methodologically we borrow tools and models of learning and evolution with local interaction developed by Ellison (1993) and Eshel, Shaked and Samuelson (1998) and later extended by Alós Ferrer and Weidenholzer (2008, 2016), or Chen, Chow and Wu (2013).

Our paper is also related to the vast literature on social norms that are

[1] Later expanded expositions can be found e.g. in Burke and Young (2011) or Binmore (2010).

sustained through community enforcement. Elinor Ostrom proposed them as a way to explain the collective solutions to social dilemmas (see, e.g. Ostrom 2000) and Cristina Bicchieri was instrumental in showing how they could be measured, as well as the importance of empirical and normative expectations from contacts (Bicchieri 2005, 2016). We add to that literature the importance of Leaders and their Followers for the establishment and survival of norms.

There is a large literature of coordination games in networks, starting from Jackson and Watts (2002) or Goyal and Vega Redondo (2005) and going to Cui (2014), Khan (2014) or Bilancini and Boncinelli (2018). Ushchev and Zenou (2020) explicitly work social norms, rather than coordination games, in a linear in means model for networks. We contribute to this literature the study of leadership and social norms.

There is also a literature in evolutionary biology (King, Johnson, Van Vugt 2009, Van Vugt, Hogan, Kaiser 2008), which considers leadership as a way for evolution to solve coordination problems. They do not consider the interaction of leader followers with crowd followers, and they do not take into account the local interaction aspect we study.

3 The model

The society consists of N players, located on a circle. Each person plays a coordination game with her k (even number) nearest neighbors in k games using a single action $x \in (A, B)$. Action A has baseline utility u_A and action B baseline utility u_B given by

$$u_x = \frac{1}{k} \sum_{j \in n_k} u\left(x, x_j\right).$$

Action A is payoff (Pareto) dominant and action B is risk dominant in the coordination game which has the following payoff matrix:

	A	B
A	d, d	e, f
B	f, e	b, b

where $d > f$, $b > e$, $d > b$, $d + e < b + f$.

In addition to the "baseline utility" from the coordination games, players experience an additional utility that depends on their type and the action of their neighbors. Each players' type is either a Leader, denoted by L; a Leader Follower, denoted by LF; or a Crowd Follower, denoted by CF.

The L player does not have a baseline utility. She can be an A supporter, meaning she has utility 1 if she uses A and 0 otherwise. The L player can also be a B supporter, with utility 1 if she uses B and 0 otherwise.

The LF player has utility $u_x + \alpha_L I_L$ when using action x. Here I_L is an indicator function taking the value 1 if she uses the action of the leader closest to her and 0 otherwise and $\alpha_L \geq 0$ reflects the charisma or influence of the leader L.

The CF player has a neighborhood of reference, comprising the k closest players on the left and right with whom she plays the coordination games. Her utility of using action x is $u_x + \alpha_C k_x/k$ where k_x is the number of her k closest neighbors using action x and $\alpha_C \geq 0$ captures the relative weight given to confirming with the reference neighborhood of their peers, "the crowd".

The L players are placed at random in the circle and their type A or B is also random. They are given a neighborhood with a fixed number of LF to their right and left, call it $l_L > k$. All players that are not LF or L are CF. Define by l_C the number of CF between two groups of LF players. We assume $l_C > 2k$.The combined assumptions of l_C and l_L is the distance between two Leaders is at least $2l_L + 2k$. We assume that this distance is an even number to avoid players with two closest leaders.

One key feature of the model is the local interaction. The coordination games are only played with the k closest neighbors in the circle; the L players affect only LF players that are "close by" and the CF are concerned whether their actions are the same as those of their neighbors. The other important distinction is between the LF players, who only get a boost by imitating the L player, and the CF, who get a boost by imitating their "peers." We believe this is a realistic feature of human interaction. We are a hierarchical species, but the group also matters to us. These concerns are probably present in different degrees in all people, but we simplify the analysis by assuming that only one

of those is relevant at a given point in time. We allow for the possibility of types shifting over time between CF and LF, if the payoff of one is clearly higher than the other, and this will play an important role in our analysis.

4 Steady states with fixed types

We first analyze the steady states of a dynamic process in which, at time $t = 0$, the LF players play the action of their closest leaders and CF players take a random action. From that period onwards, every player best responds to the actions of the players relevant to her in the previous period. The types of the players stay fixed throughout. We thus follow best-response dynamics.

The results depend on the parameter values, in particular, the baseline payoffs b, d, e, f, and the followership extra payoffs α_L.

Proposition 1 *Suppose* $d + e + 2\alpha_L < b + f - \frac{2(d-f+b-e)}{k}$. *Then everyone converges to playing* B *except the* $A-$*leaders. Suppose* $b + f > d + e + 2\alpha_L \geq b + f - \frac{2(d-f+b-e)}{k}$. *Then everyone converges to playing* B *except the* $A-$*leaders and the players between two consecutive* A *leaders, who can converge to all playing* B *or all playing* A *depending on initial conditions.*

Suppose on the other hand, that $d + e + 2\alpha_L \geq b + f$. *Then, the* LF *regions next to* A *and* B *leaders stay loyal to the leader. The region of* CF *surrounding a* B*-leader converges to playing* B *until it hits an* LF *next to an* $A-$*led region. And* CF *regions that are between to consecutive* A *leaders, can converge to all playing* B *or all playing* A *depending on initial conditions.*

Proof. We will proceed by establishing several claims that are proved in Appendix 9.1.

Claim 1 LF *with a* B *leader always follows their leader choosing strategy* B.

Claim 2 *All* CF *that are located in an area where at least one of the leaders is a* B *leader choose strategy* B.

Claims 1 and 2 establish that in steady state B leaders will be surrounded by a cluster of B play: their LF play B but so do all CF adjacent to the LF.

Hence areas that have B leaders on both sides will always turn into an all B cluster. The next claim establishes what happens if there is an A-leader on the other side of the cluster, in particular under which conditions this cluster can invade the LF region of an A leader.

Claim 3 *Any B cluster of CF can invade the LF region of an A leader till it hits the leader if $d+e+2\alpha_L < b+f$ and jump to the LF followers on the other side of the leader if $d+e+2\alpha_L < b+f-\frac{2(d-f+b-e)}{k}$ in which case all the LF of the A leader will switch to strategy B.*

The next claim confirms that the condition that the B cluster cannot invade the $A-led$ area of LF players is identical to the condition that guarantees that all LF with an A leader follow their leader choosing strategy A.

Claim 4 *If $d+e+2\alpha_L \geq b+f$ all LF with an A leader follow their leader choosing strategy A*

Claims 3 and 4 establish what happens to the LF next to an $A-led$ region. It remains to show what happens to the CFs between two A leaders.

Claim 5 *There is no stable configuration that is not a cluster of all A or all B among CFs between two A leaders*

Joining the different claims yields the proposition. ■

If the neighborhood parameter k is not too small, there are three possible outcomes in a steady state of this game. In one of them, for a sufficiently low impact of leadership α_L and sufficiently high k, only the risk dominant strategy B is capable of surviving. For a very small k the condition $d+e+2\alpha_L < b+f-\frac{2(d-f+b-e)}{k}$ cannot hold even when $\alpha_L = 0$. This condition guarantees that a B cluster of CF which invaded the LF region of an A leader (making them play B) jumps to the LF followers on the other side of the A leader. Observe, that the LF located on the other side of the A leader next to the leader has the A leader as a neighbor and a LF playing A. If this is her entire neighborhood in the coordination game, she will continue playing A since she is not affected at all by the LFs playing B on the other side of her leader.

But if her neighborhood grows, she will encounter B players; the bigger her neighborhood, the bigger her incentives to switch to playing B.

For intermediate α_L (or low α_L and sufficiently low k) there is a possibility of the Pareto dominant strategy A surviving, but this can only occur in regions between two adjacent A leaders and if the initial conditions happen to be good enough in the sense that sufficient CFs played strategy A as their initial random action. But even between two A leaders initial conditions may favor convergence to all B play. In all other regions everyone plays the risk dominant strategy B.

Finally if the leadership value is sufficiently important, then the LF players certainly play the same strategy in the coordination game as their closest leader all the time. In addition the CF players in between A leaders can converge to playing the Pareto dominant strategy A for good enough initial conditions. And all other CF players will play the risk dominant strategy B in the limit.

Note that a key aspect of this result is that once a sufficiently large cluster of agents playing one strategy or the other forms, the action happens at the boundaries of the cluster. And this is why risk dominance is so important. Someone in the boundary has half the neighbors playing one strategy and half of them playing the other. In the absence of extra elements, such as leadership, the risk dominance would take over the population.

This explains why in this result, clustering and relatively strong A leaders are crucial for the survival of the Pareto dominant, but risk dominated strategy in the limit. The risk dominant strategy does not need leadership as much, since norm following and even the pure dynamic reaction over time is sufficient to keep it in play.

5 Steady states for the evolution of types

In this section we study the evolution of types after convergence to a steady state over strategies in the coordination games has been reached. The CF can be transformed into an LF if the payoff of CF in the steady state is lower than that of CF and vice versa.

We start with a Lemma that explains under which conditions a CF does

better or worse than an LF playing the same strategy in the coordination games. It states that the extra payoff for leadership following has to be higher than the extra payoff for crowd following weighted by the proportion of people playing the strategy in the neighborhood.

Lemma 1 *A LF following her leader outperforms a CF playing the same strategy as the leader iff*

$$\alpha_L > \frac{x_k}{k}\alpha_C \tag{1}$$

where x_k are the number of neighbors playing the same strategy than the player under consideration. A LF following her leader is outperformed by a CF playing the same strategy as the leader iff (1) does not hold

Proof. Since LF and CF play the same strategy they get the same payoffs from playing the coordination games, while the LF gets additionally α_L since she follows her leader and CF gets the extra payoff from conforming to the crowd $\frac{x_k}{k}\alpha_C$. The strategy with the higher extra payoff outperforms the other strategy. ■

Now we are ready to state the main propositions of this section. For ease of expositions, we divide them according to the three relevant cases derived in Proposition 1.

Proposition 2 *Suppose $d + e + 2\alpha_L < b + f - \frac{2(d-f+b-e)}{k}$. If $\alpha_L < \alpha_C$, everybody converges to a CF playing B. If $\alpha_L > \alpha_C$ all A leaders will be surrounded by CF playing B while the $B-$leaders will be surrounded by LF playing B. In both cases everybody except the A leader plays B.*

Proof. See Appendix 9.2. ■

Recall that for α_L relatively low and k sufficiently high risk dominant play on one side of an A leader can invade the LF on other side with fixed types. Since evolution of types takes place after convergence to the steady state in strategies played in the coordination games (which leads to everybody playing B except the A leaders), LF next to an A leader clearly prefer to become CF since by playing B they don't follow their leader's strategy anyway. If in addition, α_L is low relative to α_C, the LF will do worse than the CF in

general and even LF next to a B leader with become CF. If on the other hand, α_L is high relative to α_C, then at least the LF close to A leaders play the Pareto dominant strategy. If on the other hand, α_L is high relative to α_C, then at least the LF close to B leaders will remain a LFs.

Proposition 3 *Suppose* $b + f > d + e + 2\alpha_L \geq b + f - \frac{2(d-f+b-e)}{k}$. *In this case:*

All-B regions before the evolution of types remain all-B regions after the evolution of types. Within these all−B regions everybody closest to an A-leader now is a CF *while all players closest to a B-leader are* LF *iff* $\alpha_L > \alpha_C$ *and are* CF *otherwise. For all-A regions before the evolution of types different cases apply:*

1. *they remain all-A regions after the evolution of types*

   (a) *everybody becomes* LF *playing A if*

   $$d + e + 2\alpha_L > b + f + \alpha_C \left(1 - \frac{2}{k}\right) - \frac{2\,(d - f + b - e)}{k} \quad (2)$$

   and $\alpha_L > \alpha_C$

   (b) *all former* CF *playing A remain playing A and will partially invade the* LF *playing A but not all the way up to the A−leader if (2) holds and* $\alpha_C > \alpha_L > (\frac{1}{2} + \frac{1}{k})\alpha_C$ *The first* LF *not to convert is the* LF *with the smallest y where y the number of her neighbors playing B such that by Lemma 1 condition (1) holds for* $x_k = k - y$.

   (c) *everybody becomes* CF *playing A if*

   $$\alpha_L < \left(\frac{1}{2} + \frac{1}{k}\right)\alpha_C \quad (3)$$

   and

   $$b + f < e + d + \alpha_C \frac{2}{k} + \frac{2\,(d - f + b - e)}{k} \quad (4)$$

2. *they become an all-B region after the evolution of types with only* CF *when both (2) and (4) do not hold.*

Proof. See Appendix 9.3 ■

This intermediate case for α_L is more nuanced. In the regions where everybody played B before the evolution of types , everybody keeps playing B but there may be shifts between CF and LF depending on the relative sizes of α_L and α_C. Regions where everybody played A before the evolution of types can remain all A regions or become all B regions. The latter case requires that neither α_L nor α_C are sufficiently high and converts everybody into a CF. If α_L and α_C are sufficiently high, the Pareto dominant A regions before the evolution of types are robust to the evolution of types: there can be shifts between CF and LF depending on the relative importance of α_L and α_C, but everybody will play A. If α_L is too low relative to α_C everybody becomes a CF: when α_L increases sufficiently the LFs closest to an A leader resist the CF invasion while for sufficiently high α_L the LFs will take over.

Proposition 4 *Suppose that* $d + e + 2\alpha_L \geq b + f$.

1. *Suppose that*

$$d + e + 2\alpha_L > b + f + \alpha_C \tag{5}$$

   (a) *If (5) holds and* $\alpha_L > \alpha_C$ *all players will become LF following the strategy of their closest leader.*

   (b) *If (5) holds and* $\alpha_L < \alpha_C$ *all players will play the same strategy as the closest Leader in the coordination game, but some will be CF and some will be LF, in particular:*

      i. *all players located between two leaders of the same type will be CF.*

      ii. *For players located between two leaders of different types, those furthest away from their closest leader will be LF imitating their leader while those sufficiently close to the leader will be CF. By Lemma 1 the number of neighbors* x_k *who play the same strategy in the coordination game as the player in question is defined by the lowest* x_k *for which condition (1) is violated.*

2. *If (5) does not hold so that* $b + f + \alpha_C > d + e + 2\alpha_L \geq b + f$ *all players between two leaders with different strategies will play* B *in the coordination game where everybody closest to an* A*-leader is a* CF *while all players closest to a* B *leader are* LF *iff* $\alpha_L > \alpha_C$ *and* CF *otherwise. All players between two* B*-leaders will also play* B *and are* LF *iff* $\alpha_L > \alpha_C$ *and* CF *otherwise. We have to distinguish the following cases for players between two* $A-$*leaders:*

   (a) *If before the evolution of types the* CF *converged to playing* B *those* CF *invade the area of* A*-*LF *and in the long-run everybody becomes a* CF *playing* B*.*

   (b) *For all-*A *regions between two* A *leaders before the evolution of types the results of proposition 3 apply.*

Proof. See Appendix 9.4. ■

In this last case, with α_L sufficiently large, there are more possibilities for the Pareto dominant strategy to survive.

6 The relative importance of A and B leaders

We have claimed that Leadership is less important for the long run survival of the risk dominant but Pareto dominated strategy B than for the Pareto dominant but risk dominated strategy A. To explore how much this is true, we study what happens when the charisma of leaders disappears. In particular we analyze the case where $\alpha_{L_B} = 0$ and $\alpha_{L_A} > 0$ and then check what happens for $\alpha_{L_B} \geq 0$ and $\alpha_{L_A} = 0$. Observe that even when $\alpha_{L_B} = 0$ when types are fixed all LF including the furthest away form a B leader will follow this leader choosing strategy B simply because B is risk dominant (see Claim 1). On the other hand, when types are fixed unless the A leader has a minimum charisma namely $\alpha_{L_A} > \underline{\alpha_{L_A}} = \frac{b+f-(d+e)}{2}$ (see Claim 4) the LF furthest away form this A leader might not follow this leader and deviate to the risk dominant strategy which will unravel to the LF closest to the A leader. So the steady states with fixed types as described in Proposition 1 are unaffected when

$\alpha_{L_B} = 0$ but when $\alpha_{L_A} = 0$ payoff dominant play can only occur between two A leaders when $d + e \geq b + f - \frac{2(d-f+b-e)}{k}$ and initial conditions are favorable for the CF followers. Given this result, to understand the implications of $\alpha_{L_B} = 0$ for the evolution of types, we only need to check how this assumption affects Propositions 2, 3 and 4. It is immediate from Proposition 2 that if k is sufficiently high and a_{L_A} is sufficiently low everybody will become a CF playing B since under these conditions only B leaders could have LF but now there is no value in following a B leader. Setting $\alpha_{L_B} = 0$ will eliminate the possibility to have all-B regions with LF followers in Proposition 3 for intermediate a_{L_A} but the rest of the proposition is unaffected. Similarly, the only effect of setting $\alpha_{L_B} = 0$ in Proposition 4 with $\alpha_{L_B} = 0$ is to convert LF players choosing strategy B into CF players choosing strategy B.

Summarizing, with $\alpha_{L_B} = 0$ the only difference with respect to previous results, is that in regions close to B Leaders, there will be CF players, because B Leaders now cannot attract LF. But strategy A still cannot invade regions with a B Leader: these regions will be populated by CF playing B. This is so because a CF playing B does better than a CF playing A when half her neighbors are playing B and half are playing A.While the existence of B leaders guarantees the formation of risk dominant clusters, their charisma, i.e. how attractive they are, does not matter for the choice of actions.

On the other hand since $\alpha_{L_A} = 0$ already affects the steady states with fixed types it also has important implications for the survival of all A regions when types can evolve. Introducing this assumption into Proposition 3 we learn that an all-A cluster between two A leaders can only survive the evolution of types with everybody becoming CF between the two A leaders if the payoff from rule following is sufficiently high, in particular condition (4) needs to hold which leads to

$$\alpha_C > \underline{\alpha_C} = \frac{k\,(b + f - (e + d))}{2} - (d - f + b - e)$$

The above results lead to the following remarks providing further insights.

Remark 1 *If there were only A Leaders, then for sufficiently high α_L the steady state could converge to the whole population playing A.*

Remark 2 *The risk dominant action B is always guaranteed to survive so long as there is a B Leader somewhere, while the risk dominated action disappears if α_L is low and its expansion is always limited by the existence of a B leader. In other words, the risk dominated action can never infiltrate regions where the closest Leader is a B Leader. The B Leader is a shield against infiltration no matter how charismatic A leaders are.*

6.1 Removal of Leaders

We now study another question related to Leadership. Suppose a social planner wanted to increase as much as possible the number of A players. One possibility to do this is to remove B Leaders. To avoid the simplistic case where all of them can be removed, suppose she can remove just one Leader. Which removal would lead to the largest increase in A play for fixed types?

Proposition 5 *Suppose a B Leader is taken out before the game starts, and the LF surrounding that B Leader become CFs and all other Leaders are already located, and initial conditions are random.*

If $d+e+2\alpha_L < b+f-\frac{2(d-f+b-e)}{k}$. No removal of B Leaders can make a difference in final outcome.

Suppose $d+e+2\alpha_L \geq b+f-\frac{2(d-f+b-e)}{k}$. Then, the only removal of a B Leader that can make a difference in increasing A play, is when a B Leader whose nearest Leaders on both sides are A Leaders is removed.

Proof. This is a corollary of Proposition 1 ■

The situation now is identical to the beginning of the game in general but with one less B leader. Then we know from Proposition 1 that if $d+e+2\alpha_L < b+f-\frac{2(d-f+b-e)}{k}$, everybody except A Leaders will play B in the limit so the removal of B Leaders does not make a difference. When $d+e+2\alpha_L \geq b+f-\frac{2(d-f+b-e)}{k}$, again from Proposition 1, there can be clusters of A players in between two A Leaders, hence the only change can occur if the extirpated B Leader is in between two A Leaders.

By Proposition 1, for sufficiently high α_L, CF regions between two consecutive A leaders can converge to either playing all A or playing all B depending

on initial conditions. The following proposition explains how the distance between these leaders affects the probability of convergence to one or the other strategy.

Proposition 6 *From a random initial condition between two A Leaders, if the distance between them becomes large enough, then all CF converge to playing B.*

Proof. If a k-cluster of B players form, all players will end up playing B. The chance of a k-player cluster forming at random at $t = 0$ increases as the distance between two A Leaders grows. ■

Hence, a social planner who can only remove one B leader surrounded by the two A leaders faces a trade-off between increasing the ex ante probability of reaching an A cluster (this probability is maximized by removing a B leader between to A leaders that have the shortest distance among them) and increasing the size of the A cluster should it be reached (this size is maximized by removing a B leader between two A leaders that have the biggest distance among them).

Suppose, on the other hand, that one B Leader can be removed after play in the game with fixed types has reached the steady state. The first round after the removal, the LF types stay as LF, and the CF players stay as CF but they reoptimize the strategy in the coordination game .

Proposition 7 *Suppose* $d+e+2\alpha_L < b+f$. *No removal of B Leaders after the game with fixed types has reached the steady state can make a difference in final outcome.*

Proof. When $d+e+2\alpha_L < b+f$ we are either in the context that everybody converged to playing B except for players between two consecutive A leaders with favorable initial conditions in case $d + e + 2\alpha_L > b + f - \frac{2(d-f+b-e)}{k}$ before any leader is removed. By Propositions 2 and 3 no new all A regions can evolve, and some might be preserved if we already have an all A region between two A leaders before the evolution of types, so removing a B leader will not create a new all A region. ■

Proposition 8 *Suppose $b + f < d + e + 2\alpha_L$. The removal of a B leader after the game with fixed types has reached the steady state will only make a difference in final outcome when $d + e + 2\alpha_L > b + f + \alpha_C$ implying that (5) holds and at least one of the closest leaders of the removed B leader is an A leader.*

1. *If the removed B leader was located between two A leaders, all people under the former influence of this B leader will play A.*

2. *If the removed B leader was located between an A and a B leader, A play will grow in the new area of influence of the A leader.*

3. *The best candidate for removal of a B leader located between two A leaders is the one with the greatest distance among these two A leaders.*

4. *The best candidate for removal of a B between an A leader and a B leader is the one resulting in the biggest new area of influence of an A leader who was the unique closest A leader of the removed B leader.*

5. *The overall gain in A play is greatest by removing the B leader between two A leader if the area of influence of this B leader is bigger than the biggest new area of influence of an A leader who was the unique closest A leader of the removed B leader. Otherwise the latter should be removed.*

Proof. If condition (5) does not hold no new A clusters can be created, since the B-CF of the removed B leader will invade any $A - LF$ and there is no difference in final outcome.

Clearly eliminating a B leader that is located between two B leaders will never make a difference.

When $b+f < d+e+2\alpha_L$ all LF with an A leader follow their leader choosing strategy A before the types are allowed to change. So, if we eliminate a B leader between two A leaders all the LF of that eliminated B leader will now have as their closest leader an A leader and play A. This proves (1). Similarly, if the eliminated B leader is located between an A and a B leader, the LF of the eliminated leader closest to the A leader will play strategy A while those closest to the B leader will play strategy B in the first round after the removal.

All the CF will continue playing B since B is risk-dominant and at least half of their neighbors play B. We have shown in the proof of Proposition 4 that if (5) holds when the possibility to change types holds, $A-LF$ will invade the $B-CB$ regions and everybody will play the same strategy in the coordination game as their closest leader. Therefore if a B leader between two A leaders is removed the entire influence area of the removed B leader turns into playing A while if the removed B leader only had one closest A leader A play grows in the new influence area of this A leader. This proves (2). Then (3), (4) and (5) are straightforward implications of (1), and (2). ■

Finally, assume the removal of one B Leader happens after a steady state in the evolution of types has been reached.

Proposition 9 *Suppose $b+f < d+e+2\alpha_L$. The removal of a B leader after the steady state in the evolution of types has been reached only makes a difference in final outcome when $d+e+2\alpha_L > b+f+\alpha_C$ implying that (5) holds and at least one of the closest leaders of the removed B leader is an A leader.*

1. *If the removed B leader was located between two A leaders, all people under the former influence of this B leader will play A.*

2. *If the removed B leader was located between an A and a B leader, A play will grow in the new area of influence of the A leader.*

Proof. Clearly eliminating a B leader that is located between to B leaders will never make a difference.

Assume $d+e+2\alpha_L > b+f+\alpha_C$.When a B Leader between two A Leaders is removed, in the first round, the B LF around the removed B Leader become A LF. By Proposition 4 if $\alpha_L > \alpha_C$ this is the final outcome. If $\alpha_L < \alpha_C$, then in the first round, the B CF around the removed B Leader stay B CF, because at most half of their neighbors play A. From the next round onwards, the B CF start being invaded by the A LF, while on the other side the A LF is is invaded by the A CF that border the A LF. At the end, though, everybody will be A CF, hence everybody formerly under the influence of the removed B leader will be an added A player. This proves (1).

When a B Leader between an A and a B Leader is removed, in the first round, only the B LF around the removed B Leader who fall into the new area of influence of the A leader become A LF. If $\alpha_L > \alpha_C$ everybody in the new influence area of the A leader becomes A LF and this is the final outcome. If $\alpha_L < \alpha_C$ then by Proposition 4 the former B LF that become A LF in the first round are those located furthest away from the removed B leader and hence bordering the A LF area before the removal of the B leader. The B-CF now under the influence of the A leader continue playing B since B is risk-dominant and at least half of their neighbors play B. Since (5) holds, A-LF can invade these $C - LF$ while on the other side the A LF is is invaded by the A CF that border the A LF. At the end, though, everybody under the influence of the A leader will play A with those closest to the A leader being CF and those furthest away being LF, so everybody in the new influence area of the A leader will be a new A player. This proves (2). ■

Observe that independently of the timing of the removal of the B leader, the removal of a B leader between an A Leader might enhance all A play.If A Leaders are sufficiently charismatic (α_L is large), then this is guaranteed if the removal happens either after the steady state in strategies is reached with fixed types, or after the steady state of the evolution of types is reached. In both cases, the biggest impact happens if the B Leader that is removed is located between the two A Leaders that are furthest apart. If the removal of the B Leader happens at the beginning of the game, it depends on initial conditions whether A play is enhanced. This is more likely the closer the two consecutive A Leaders surrounding the eliminated B Leader are located. Of course, at the same time, if they are close to one another, the number of affected players is smaller.

If A Leaders are sufficiently charismatic (α_L is large) A play also grows if a B-leader located between an A-leader and a B-leader is removed either after the steady state in strategies is reached with fixed types, or after the steady state of the evolution of types is reached. The area of influence of the neighboring A-leader will grow and everybody in this area will play the Pareto dominant action, hence the size of the growth of Pareto dominant play corresponds to the size of the increase in the area of influence of this

neighboring $A-$leader.

7 Conclusion

We have postulated a game in which leadership and the following of norms interact, in an environment where individuals play a coordination game with local interaction. We find that the survival of Pareto efficient outcomes over time is very dependent on clustering and on the existence and strength of Leaders willing to support the actions leading to that outcome.

There are several important extensions that could be considered for this model. We assume that people either follow the *L*eaders, or follow their peers. Mixed motivations could be important. The extent of peer influence is limited to a small environment, which is consistent with the evidence about the cognitive limitation on human relationships (the Dunbar numbers, see e.g. Dunbar 1992 and Dunbar and Shultz 2007). But we have focused on a particularly simple network structure, where the evolution is relatively tractable. More complex structures might produce interesting results. In particular, Leaders that can reach different sizes of the population seem a worthwhile avenue for future research.

8 References

Acemoglu, Daron, and Matthew O. Jackson. "History, expectations, and leadership in the evolution of social norms." The Review of Economic Studies 82.2 (2015): 423-456.

Acemoglu, Daron, and Matthew O. Jackson. "Social norms and the enforcement of laws." Journal of the European Economic Association 15.2 (2017): 245-295.

Alós-Ferrer, Carlos, and Simon Weidenholzer. "Contagion and efficiency." Journal of Economic Theory 143.1 (2008): 251-274.

Alós-Ferrer, Carlos, and Simon Weidenholzer. "Imitation, local interactions, and efficiency." Economics Letters 93.2 (2006): 163-168.

Ang, James B., Per G. Fredriksson, and Swati Sharma. "Individualism and the adoption of clean energy technology." Resource and Energy Economics 61 (2020): 101180.

Belk, Russell and Gülnur Tumbat (2005) The Cult of Macintosh, Consumption Markets & Culture, 8:3, 205-217

Bicchieri, Cristina. The grammar of society: The nature and dynamics of social norms. Cambridge University Press, 2005.

Bicchieri, Cristina. Norms in the wild: How to diagnose, measure, and change social norms. Oxford University Press, 2016.

Bilancini, Ennio, and Leonardo Boncinelli. "Social coordination with locally observable types." Economic Theory 65.4 (2018): 975-1009.

Binmore, Ken, and Larry Samuelson. "An economist's perspective on the evolution of norms." Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft 150.1 (1994): 45-63.

Binmore, Ken. "Social norms or social preferences?." Mind & Society 9.2 (2010): 139-157.Burke, Mary A., and H. Peyton Young. "Social norms." Handbook of social economics. Vol. 1. North-Holland, 2011. 311-338.

Chen, Hsiao-Chi, Yunshyong Chow, and Li-Chau Wu. "Imitation, local interaction, and coordination." International Journal of Game Theory 42.4 (2013): 1041-1057.

Cui, Zhiwei. "More neighbors, more efficiency." Journal of Economic Dynamics and Control 40 (2014): 103-115.

Cummins, Denise. "Dominance, status, and social hierarchies." The handbook of evolutionary psychology (2005): 676-697.

Dunbar, Robin IM. "Neocortex size as a constraint on group size in primates." Journal of human evolution 22.6 (1992): 469-493.

Dunbar, Robin IM, and Susanne Shultz. "Evolution in the social brain." science 317.5843 (2007): 1344-1347.

Ellison, Glenn. "Learning, local interaction, and coordination." Econometrica: Journal of the Econometric Society 61.5 (1993): 1047-1071.

Eshel, Ilan, Larry Samuelson, and Avner Shaked. "Altruists, egoists, and hooligans in a local interaction model." American Economic Review 88.1 (1998): 157-179.

Fang, Tommy Pan, Andy Wu, and David R. Clough. "Platform diffusion at temporary gatherings: Social coordination and ecosystem emergence." Strategic Management Journal 42.2 (2021): 233-272.

Farrell, Joseph, and Garth Saloner. "Coordination through committees and markets." The RAND Journal of Economics (1988): 235-252.

Gilbert, Paul. "Evolution and social anxiety: The role of attraction, social competition, and social hierarchies." Psychiatric Clinics 24.4 (2001): 723-751.

Goyal, Sanjeev, and Fernando Vega-Redondo. "Network formation and social coordination." Games and Economic Behavior 50.2 (2005): 178-207.

Holmberg, Tove, Marcus Logander, and Fredrik Lindqvist. "" Living on the Edge"-A Case Study of Important Factors for the Survival of Apple Computers, Inc." (2005).

Iriberri, Nagore, and José-Ramón Uriarte. "Minority language and the stability of bilingual equilibria." Rationality and Society 24.4 (2012): 442-462.

Jackson, Matthew O., and Alison Watts. "On the formation of interaction networks in social coordination games." Games and Economic Behavior 41.2 (2002): 265-291.

Khan, Abhimanyu. "Coordination under global random interaction and local imitation." International Journal of Game Theory 43.4 (2014): 721-745.

King, Andrew J., Dominic DP Johnson, and Mark Van Vugt. "The origins and evolution of leadership." Current biology 19.19 (2009): R911-R916.

Lempert, Robert J., Alan H. Sanstad, and Michael E. Schlesinger. "Multiple equilibria in a stochastic implementation of DICE with abrupt climate change." Energy economics 28.5-6 (2006): 677-689.

Levy, Steven. 2000. Insanely great: The life and times of Macintosh, the computer that changed everything. New York: Penguin.

Maner, Jon K., and Charleen R. Case. "Dominance and prestige: Dual strategies for navigating social hierarchies." Advances in experimental social psychology. Vol. 54. Academic Press, 2016. 129-180.

Müller, Malte, Christian Kimmich, and Jens Rommel. "Farmers' adoption of irrigation technologies: experimental evidence from a coordination game with positive network externalities in India." German Economic Review 19.2 (2018): 119-139.

Ostrom, Elinor. "Collective action and the evolution of social norms." Journal of economic perspectives 14.3 (2000): 137-158.

Wu, Ruijuan, Cheng Lu Wang, and Andy Hao. "What makes a fan a fan?: The connection between Steve Jobs and Apple fandom." Handbook of research on the impact of fandom in society and consumerism. IGI Global, 2020. 378-396.

Sunstein, Cass R. "On academic fads and fashions." Mich. L. Rev. 99 (2000): 1251.

Ushchev, Philip, and Yves Zenou. "Social norms in networks." Journal of Economic Theory 185 (2020): 104969.

Van Vugt, Mark, Robert Hogan, and Robert B. Kaiser. "Leadership, followership, and evolution: some lessons from the past." American Psychologist 63.3 (2008): 182.

Venkatraman, N., and Chi-Hyon Lee. "Preferential linkage and network evolution: A conceptual model and empirical test in the US video game sector." Academy of Management Journal 47.6 (2004): 876-892.

Young, H. Peyton. 1993. The evolution of conventions. Econometrica 61, 57-84.

Young, H. Peyton. "Social norms and economic welfare." European Economic Review 42.3-5 (1998): 821-830.

9 Appendix

9.1 Proof of the Claims leading to Proposition 1.

Claim 1: LF with a B leader always follows their leader choosing strategy B.

Proof. It suffices to look at the most distant LF to her B leader who has a payoff at least $\frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)+\alpha_L$ choosing B and at most $\frac{1}{k}\left(d\frac{k}{2}+e\frac{k}{2}\right)$ choosing A, so from $b+f>d+e$ strategy B is the best response. ■

Claim 2: All CF that are located in an area where at least one of the leaders is a B leader choose strategy B.

Proof. Take a CF that is located next to a B-led region where all LF play

B by Claim 1. Her payoff from playing B is at least $\frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)+\alpha_C\frac{1}{2}$. Her payoff from playing A is at most $\frac{1}{k}\left(d\frac{k}{2}+e\frac{k}{2}\right)+\alpha_C\frac{1}{2}$. Hence she will choose B because $b+f>d+e$. By induction, all the CF next to a B-led region flip to B. ■

Claim 3: Any B cluster of CF can invade the LF region of an A leader till it hits the leader if $d+e+2\alpha_L<b+f$ and jump to the LF followers on the other side of the leader if $d+e+2\alpha_L<b+f-\frac{2(d-f+b-e)}{k}$ in which case all the LF of the A leader will switch to strategy B.

Proof. We know that the most distant LF to her A leader facing a B cluster invasion (from the left) has a payoff $\frac{1}{k}\left(d\frac{k}{2}+e\frac{k}{2}\right)+\alpha_L$ choosing A $\frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)$ choosing B, so from $d+e+2\alpha_L<b+f$ she flips to playing B. By induction this frontier keeps advancing until it hits the A leader. Now the LF to the right of the A leader has a payoff $\frac{d}{k}\left(\frac{k}{2}+1\right)+\frac{e}{k}\left(\frac{k}{2}-1\right)+\alpha_L$. Her payoff from playing B is $\frac{b}{k}\left(\frac{k}{2}-1\right)+\frac{f}{k}\left(\frac{k}{2}+1\right)$ so she flips if

$$d+e+2\alpha_L<b+f-\frac{2\left(d-f+b-e\right)}{k}$$

■

Claim 4: If $d+e+2\alpha_L\geq b+f$ all LF with an A leader follow their leader choosing strategy A.

Proof. Again it suffices to look at the most most distant LF to the A leader. She has a payoff of at least $\frac{1}{k}\left(d\frac{k}{2}+e\frac{k}{2}\right)+\alpha_L$ choosing A and of at most $\frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)$ choosing B, so from $d+e+2\alpha_L\geq b+f$ she stays playing A. ■

Claim 5: There is no stable configuration that is not a cluster of all A or all B among CFs between two A leaders

Proof. Take a CF person that is in a sector with x A neighbors and $k-x$ B neighbors. The payoff of A is $\frac{xd+(k-x)e}{k}+\alpha_C\frac{x}{k}$. The payoff from B is $\frac{xb+(k-x)f}{k}+\alpha_C\frac{k-x}{k}$. A is better than B if

$$\frac{x}{k}>\frac{(f-e)+\alpha_C}{(d+f)-(e+b)+2\alpha_C} \tag{6}$$

Suppose 6 holds for an A sitting next to a B to the left of B. Then we will

show B wants to flip to A. Observe that the difference in the neighborhood between A and B is that there is one person to the extreme left of A interval, call it C that does not belong $B's$ interval and one person to the extreme right of $B's$ interval that does not belong to A interval, call it D, and A has B as a neighbor and B has A as a neighbor. Assume first that 6 holds.

Case 1. C is A and D is A.Then B has one more A neighbor than A, so B wants to switch to A.

Case 2. C is A and D is B. Then B has same amount of A neighbors than A, so B wants to switch to A.

Case 3. C is B and D is A. Then B has two more A neighbors than A, so B wants to switch to A.

Case 4. C is B and D is B. Then B has one more A neighbors than A, so B wants to switch to A.By induction this unravels to all CF players between two A leaders.

Suppose 6 does not holds for an A sitting next to a B to the left of B. Then a analogous argument shows that A wants to flip to B. By induction this unravels to all CF players between two A leaders. ■

9.2 Proof of Proposition 2

Suppose $d+e+2\alpha_L < b+f-\frac{2(d-f+b-e)}{k}$ implying that everybody except for the $A-$leaders chooses strategy B before the evolution of types.Then the LF next to an A leader who played B will switch to a CF playing B because she does not get any benefit from following the leader but gets benefits from following the crowd that all play the same strategy. For the LF and CF closest to a B leader, they choose to be LF by Lemma (1) iff $\alpha_L > \alpha_C$ since $x_k = k$ because all neighbors play B.

9.3 Proof of Proposition 3

Suppose $b+f > d+e+2\alpha_L \geq b+f-\frac{2(d-f+b-e)}{k}$. In this case, all the regions playing B stay playing B but the CF playing B will invade the LF who play B in regions next to an A leader because these LF don't follow their leader

and choose B, so they are better following the crowd. All the players in an B region that are closest to a B leader will become LF iff $\alpha_L > \alpha_C$ since they are surrounded by only all$-B$ neighbors and becomes CF iff $\alpha_L < \alpha_C$.

Now we study what happens to the regions between two $A-$leaders that converged to playing A before the evolution of types sets in.

Consider the first LF playing A next to an A Leader. She prefers staying LF playing A instead of switching to a CF playing B if

$$\begin{aligned}
\frac{d}{k}\left(\frac{k}{2}+1\right)+\frac{e}{k}\left(\frac{k}{2}-1\right)+\alpha_L &> \frac{b}{k}\left(\frac{k}{2}-1\right)+\frac{f}{k}\left(\frac{k}{2}+1\right)+\frac{\alpha_C}{k}\left(\frac{k}{2}-1\right)\\
d\left(\frac{1}{2}+\frac{1}{k}\right)+e\left(\frac{1}{2}-\frac{1}{k}\right)+\alpha_L &> b\left(\frac{1}{2}-\frac{1}{k}\right)+f\left(\frac{1}{2}+\frac{1}{k}\right)+\alpha_C\left(\frac{1}{2}-\frac{1}{k}\right)\\
d+e+2\alpha_L &> b+f+\alpha_C\left(1-\frac{2}{k}\right)-\frac{2\left(d-f+b-e\right)}{k}
\end{aligned}$$

She prefers switching to CF playing A instead of staying LF playing A

$$\begin{aligned}
\frac{d}{k}\left(\frac{k}{2}+1\right)+\frac{e}{k}\left(\frac{k}{2}-1\right)+\alpha_L &< \frac{d}{k}\left(\frac{k}{2}+1\right)+\frac{e}{k}\left(\frac{k}{2}-1\right)+\frac{\alpha_C}{k}\left(\frac{k}{2}+1\right)\\
2\alpha_L &< \alpha_C\left(1+\frac{2}{k}\right)
\end{aligned}$$

which is equivalent to (3).

So the condition to stay a LF is

$$2\alpha_L > \max\{\alpha_C\left(1+\frac{2}{k}\right), b+f-(d+e)+\alpha_C\left(1-\frac{2}{k}\right)-\frac{2\left(d-f+b-e\right)}{k}\} \tag{7}$$

Notice that for that person being a CF playing B is worse than a CF playing A if

$$\begin{aligned}
\frac{b}{k}\left(\frac{k}{2}-1\right)+\frac{f}{k}\left(\frac{k}{2}+1\right)+\frac{\alpha_C}{k}\left(\frac{k}{2}-1\right) &< \frac{d}{k}\left(\frac{k}{2}+1\right)+\frac{e}{k}\left(\frac{k}{2}-1\right)+\frac{\alpha_C}{k}\left(\frac{k}{2}+1\right)\\
b\left(\frac{1}{2}-\frac{1}{k}\right)+f\left(\frac{1}{2}+\frac{1}{k}\right)+\alpha_C\left(\frac{1}{2}-\frac{1}{k}\right) &< d\left(\frac{1}{2}+\frac{1}{k}\right)+e\left(\frac{1}{2}-\frac{1}{k}\right)+\alpha_C\left(\frac{1}{2}+\frac{1}{k}\right)\\
b+f &< e+d+\alpha_C\frac{2}{k}+\frac{2\left(d-f+b-e\right)}{k}
\end{aligned}$$

which is (4).

1. Assume (2) holds and $\alpha_L > \alpha_C$ which implies that (3) is violated. Then everybody between two A Leaders becomes an LF playing A since LF-A dominates $CF - A$ even when all neighbors play A.

2. Assume (2) holds and (3) is violated. Moreover, $\alpha_L < \alpha_C$ so that combined with (3) violated the parameter restriction becomes $\alpha_C > \alpha_L > (\frac{1}{2} + \frac{1}{k})\alpha_C$. In this case the CF playing A invade the LF regions but do not take it over completely. The CF invasion stops at the biggest distance $k - y$ from the A (where y is the number of neighbors on the B region of the A Leader.) satisfying $\alpha_L > \frac{k-y}{k}\alpha_C$ which guarantees that condition (1) of Lemma 1 is satisfied.

3. Assume (3) holds and (4) holds. The CF playing A dominates both CF playing B and LF playing A, so everybody will become CF playing A

4. Assume (2) and (4) are both violated. Then CF playing B dominates both CF playing A and LF playing B, so the former all A region becomes and all-B regions with all players between the A Leaders becoming CF and play B

9.4 Proof of Proposition 4

Look at the B region boundary with an A LF who has to decide to switch to CF playing B (she cannot switch to CF playing A because with half of the neighborhood playing B a CF always plays B). That person stays LF if

$$\begin{aligned} \frac{1}{k}\left(d\frac{k}{2} + e\frac{k}{2}\right) + \alpha_L &> \frac{1}{k}\left(b\frac{k}{2} + f\frac{k}{2}\right) + \alpha_C\frac{1}{2} \\ d + e + 2\alpha_L &> b + f + \alpha_C \end{aligned}$$

which coincides with (5). Under (5) all the CF players playing B next to an A LF playing A decide to switch to A playing LF as long as their closest leader is A since they face exactly the trade-off described to derive (5). When the closest leader becomes B then the CF playing B have to decide whether to

become a LF playing B. The first one who is surrounded on one side by all B players and on the other by all-A players, she compares

$$\begin{aligned} \frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)+\alpha_L &> \frac{1}{k}\left(b\frac{k}{2}+f\frac{k}{2}\right)+\alpha_C\frac{1}{2} \\ b+f+2\alpha_L &> b+f+\alpha_C \end{aligned}$$

which definitely holds because $b+f>d+e$ and (5) holds. The next player compares

$$\frac{1}{k}\left(b\frac{k+1}{2}+f\frac{k-1}{2}\right)+\alpha_L > \frac{1}{k}\left(b\frac{k+1}{2}+f\frac{k-1}{2}\right)+\alpha_C\frac{k/2+1}{k}$$

so the frontier keeps advancing until $\frac{x_k}{k} < \frac{a_L}{\alpha_C} = d^* > \frac{1}{2}$ so that condition (1) of Lemma 1 is violated.

If $\frac{a_L}{\alpha_C} = d^* > 1$ this will never happen and then all CF playing B closest to a B leaders will become LF playing B.

Proof. If (5) holds and $\alpha_L > \alpha_C$ or equivalently $d^* > 1$ the A LF will advance as long they are closest to an A leader and everybody closest to a B leader becomes an LF playing B. All areas between two A leaders are converted two A-LF who either invade the former $B-CF$ area (since (5) holds) or the former $A-CF$ area (since $\alpha_L > \alpha_C$) located in the middle between these two A leaders. All players located between two B leaders are surrounded by only B neighbors and will become $B-LF$ since $\alpha_L > \alpha_C$.

If (5) holds and $\alpha_L < \alpha_C$ or equivalently $d^* < 1$ then at the frontier of and A LF area with an B-CF area the A LF will invade the neighboring $B-CF$ players as long they are closest to an A leader and players closest to a B leader play B. Those with the closest B-leader closest to the frontier of the all A area so that $\frac{x_k}{k} > d^*$ are LF playing B. And all the others with a closest B leader are CF playing B. Everybody located between two leaders of the same type will become a CF playing the strategy of its nearest leader since $\alpha_L < \alpha_C$ and the leader is always surrounded by players playing the same strategy as the leader in the coordination game.

If (5) does not hold, then the A LF switch to B CF until we hit the A Leader. From that point on, the analysis that we did in the previous proposi-

tion holds. ■