Ca' Foscari University of Venice

Department of Economics

Working Paper

Sebastiano Della Lena

Pietro Dindo

On the Evolution of Norms in Strategic Environments

ISSN: 1827-3580

No. 16/WP/2019

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 16/WP/2019
ISSN 1827-3580

On the Evolution of Norms in Strategic Environments

Sebastiano Della Lena
Ca' Foscari University of Venice

Pietro Dindo
Ca' Foscari University of Venice

Abstract
In a heterogeneous population divided into two cultural groups, we investigate the intergenerational dynamics of norms, modeled as preferences over actions, as depending on strategic environments. We find that environments with strategic complementarity or substitutability lead to different long-run norms and horizontal socializations. When players face many games within the same class, under complementarity agents converge to the same norm and socialization is high, under substitutability norms may diverge or become neutral and socialization is low. However, for specific games, partial convergence can arise under complementarity, providing an explanation to cultural heterogeneity, and partial divergence can arise under substitutability.

Keywords
Evolution of Norms, Cultural Transmission, Endogenous preferences, Cultural Heterogeneity

JEL Codes
C7, D9, I20, J15, Z1

Address for correspondence:
Sebastiano Della Lena
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: sebastiano.dellalena@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/pag/16882/)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

On the Evolution of Norms in Strategic Environments*

Sebastiano Della Lena † Pietro Dindo‡

June 17, 2019

Abstract

In a heterogeneous population divided into two cultural groups, we investigate the intergenerational dynamics of norms, modeled as preferences over actions, as depending on strategic environments. We find that environments with strategic complementarity or substitutability lead to different long-run norms and horizontal socializations. When players face many games within the same class, under complementarity agents converge to the same norm and socialization is high, under substitutability norms may diverge or become neutral and socialization is low. However, for specific games, partial convergence can arise under complementarity, providing an explanation to cultural heterogeneity, and partial divergence can arise under substitutability.

Journal of Economic Literature Classification Numbers: C7, D9, I20, J15, Z1

Keywords: Evolution of Norms, Cultural Transmission, Endogenous preferences, Cultural Heterogeneity.

*We thank Alberto Bisin, Sergio Currarini, Ignacio Monzon, Fabrizio Panebianco, Paolo Pin, William H. Sandholm, Lambert Schoonbeek, Adriaan Soetevent, Simon Weidenholzer, Thierry Verdier, and seminar a and conferences participants at Groningen University, RES 2018, RES 2018 Junior Symposium, 10th VPDE PhD Workshop in Economics, 12th BiGSEM Doctoral Workshop on Economic Theory, ASSET 2017 for helpful comments and stimulating discussions. All errors and mistakes remain ours.

†*Corresponding Author*: Department of Economics, Ca' Foscari University of Venice. E-mail: sebastiano.dellalena@unive.it

‡Department of Economics, Ca' Foscari University of Venice. E-mail: pietro.dindo@unive.it

From any given set of rules of conduct of the element will arise a steady structure (showing 'homeostatic' control) only in an environment in which there prevails a certain probability of encountering the sort of circumstances to which the rules of conduct are adapted. A change of environment may require, if the whole is to persist, a change in the order of the group and therefore in the rules of conduct of the individuals; and a spontaneous change of the rules of individual conduct and of the resulting order may enable the group to persist in circumstances which, without such change, would have led to its destruction.[Hayek 1967: 71]

1 Introduction

In our societies some norms or cultural traits, for example languages, appear to be more homogeneous than others, like attitudes toward conflict or effort choices at the workplace. This might be due to the fact that material incentives to coordinate are, often, stronger in the former strategic environment than in the latter. In fact, in linguistic interactions there are evident material incentives to coordinate on the same language, while in competitive interactions for scarce shared resources, the material incentives are to anti-coordinate. Nevertheless, cultural heterogeneity is ubiquitous in many societies even when there are strong incentives to coordinate. For example, we observe separated minorities who fail to use incumbents' languages and the resilience of the native languages in integrated second-generation immigrants.[1]

The aim of the paper is to propose a model of the interplay between norms and strategic decisions that allows to study the evolution of norms depending on the underlying strategic environment, and to explore its properties for policy purposes. The final objective is to obtain different social outcomes, such as convergence toward the same social norm, the persistence of norms' heterogeneity, or even polarization of norms, while using the same norm formation model and depending on the strategic environment agents are exposed to during their adult life. In particular, we consider environments where actions are strategic **complements**, so that agents have clear material incentives to coordinate, and environments where actions are strategic **substitutes**, so that agents have clear material incentives to anti-coordinate.[2] We explore the effect of the strategic environment on norms' evolution along two dimensions. On the one hand, we consider the effect of two strategic classes, complements or substitutes environments. On the other hand, we study

[1] Bisin and Verdier (2011) offers a review of empirical examples of cultural heterogeneity and resilience of cultural traits. For example, the slow rate of immigrants' integration in Europe and US, the persistence of 'ethnic capital' in second- and third-generation immigrants, minorities' strong attachment to languages and cultural traits.

[2] There exist strategic environments, like the prisoner's dilemma, who do not belong to either class. Our model can be applied also to these cases but it is beyond the scope of this paper.

both situations in which agents play several games with payoffs distributed within the same strategic class, and also environments in which they always play the same game. The model should help to investigate whether strategic complements (arising for example in communication problems) necessarily lead to assimilation of norms or, instead, leave space to a multicultural society or even to the arising of an oppositional culture with the consequent separation of the minority. Relatedly, we shall use the model to investigate whether in environments with strategic substitutability, being more "competitive", norms are doomed not to be homogeneous (leading for example to different attitudes toward conflict). Moreover, we wonder if there exist socio-economic environments where material incentives make norms neutral in the long run. Understanding the relationship between the strategic environment and the emergence of different cultural traits may help to better address policy issues.

In our model the population is divided into two groups or communities. Agents belonging to the same community are endowed with the same cultural trait or personal norm. Norms can be viewed as "mental representations of appropriate behavior" (Aarts and Dijksterhuis, 2003) or "internal standard of conduct" (Schwartz, 1977), namely they represent preferences over actions.[3] Agents interact twice during their lives. First, while young, each agent forms a new norm taking into account both family pressure (*vertical socialization*) and peer pressure (*horizontal socialization*).[4] The latter represents the willingness to conform to peers. Then, in the adult age, agents are randomly matched to play symmetric 2×2 games where, by shaping agents' preferences, different norms lead to different strategic outcomes (as in Akerlof, 1976; Young, 1998). Norms parametrize agent preferences over material payoffs and thus best reply actions and Nash equilibria depend both on a material component and on an immaterial one. The material component pushes toward coordination or anti-coordination depending on the ordering of two material forces that define the game class. The immaterial component can be neutral (no effect) or enhance payoffs received by playing either action. The tension between the material and immaterial forces determines ultimate payoffs and thus Nash Equilibria. In turn, actions most played in equilibrium modify each group norm and socialization level, thus different socio-economic environments can have an effect on the selection of norms as suggested by Hayek (1967). At the end of their lives (old age), each agent transmits a norm and a socialization parameter to her offsprings and the whole process repeats again. Notably, in this model agents are active in all the stages of their life (youth, adult age, and old age) and each stage can be analyzed separately.

In Section 2.1, we analyze the norm formation mechanism of the youth age. Children are not passive during the transmission process and they are responsible for the formation

[3] This definition of norms differs from the one used in the evolutionary game theory literature (Kandori, 1992; Young, 1993, for example) where a norm is broadly defined as a specific equilibrium of a strategic interaction.

[4] We refer to Cavalli-Sforza and Feldman (1981); Bisin and Verdier (2001) for the terminology.

of their own norms. We follow Kuran and Sandholm (2008) and assume that, in choosing the norm, each agent faces a trade-off between her inherited norm and a social coordination payoff, minimizing a loss function. The socialization parameter, which determines the weight put on the entire population average norm (horizontal socialization) as opposed to the weight put on the inherited norm (vertical socialization) describes the strength of such trade-off. With respect to the previous literature, we introduce heterogeneity in the socialization parameter of each community, an endogenous variable of our model.

In Section 2.2, we model the effect of norms on different strategic environments. We interpret each agent's norm as the preference for playing a particular action, thus modifying the payoff associated with playing according to it. The intuition is that the less an action is in line with the personal norm of the agent, the lower is the psychological utility derived from its associated material outcome. Each strategic environment is a symmetric 2×2 game, which is meant to be representative of tasks that people can face in their adult age. For the complements case, a possible example is the choice of which language to use when there are material incentives associated to coordination and each group norm represents the preference for using a specific language. For the substitutes case, a possible example is the choice to "fight or flight" in a competition for a shared resource, so that the material incentives are to anti-coordinate, and each group norm represents the attitude toward aggressive behavior.[5] In their adult age, agents can face several games of the same strategic class, described by a distribution of payoffs. Agents are randomly matched so that each agent strategically interacts with agents belonging to both groups. We assume a multiplicative interaction between norms and material payoffs. If the norms are neutral, the payoffs of the game are equivalent to material payoffs and agents play the original 2×2 game. If norms assume extreme values, agents stick to the associated action, giving no importance to material payoffs. When norms have intermediate values, there is a trade-off between the material consequence of actions and playing according to the behavior associated with such norm. Relatedly, we derive the possible Nash equilibria as depending on the tension between material payoffs and norms over behavior.

In Section 2.3, in order to characterize the feedback between the strategic environment and norms, we study the transmission of norms from old to young and the evolution of the socialization parameter. The transmission is moved by **cognitive dissonance** and **cultural substitution**. Cognitive dissonance, firstly proposed by Festinger (1962) and assumed also in the cultural dynamics model of Kuran and Sandholm (2008), is the tendency of agents to have consistency between behavior (actions) and norms (preferences over actions) and it is widely documented in social psychology (Cooper and Fazio, 1984; Baumeister, 1982). Cultural substitution captures the idea that the vertical socialization

[5] There exist examples of norms that can fit in both complements and substitution environments. Consider the choice of effort level in team-work: whether the underlying material payoffs exert complementarity or substitutability depends on the possibility to free-ride on the team-mate's effort. See also the discussion in Session 4.

level of offsprings negatively depends on the diffusion of parents' behavior in the population (Bisin and Verdier, 2001). The functional form chosen for the evolution of the socialization parameter is microfounded in Appendix B.

In Section 3, we derive all possible long-run outcomes as depending on both the youth coordination game and the adult-age strategic environment. We consider two limit cases: uniform and point distribution of one-stage game material payoffs. In Section 3.1, we analyze uniform distribution of payoffs and show that environments with complements or substitutes produce very different norms in the long-run. In a social environment with complements, cultural assimilation emerges as a stable steady state, i.e. both communities share the same norm and behavior in the long-run. On the contrary, in a social environment with substitutes, steady states with assimilation exist but are unstable. Since agents play a coordination game in their youth, when they face strategic substitutes in the adult age two opposite forces are at play, thus material incentives may lead both to the erosion of norms or to the polarization of norms and behavior of agents belonging to different communities. Steady states with norms' erosion or polarization may be both locally or globally stable depending on the relative size of cognitive dissonance and horizontal socialization: a higher cognitive dissonance, with respect to the maximum of horizontal socialization, corresponds to a wider space for the polarization of norms.

The different strategic environments lead also to different socialization levels. Under assimilation, agents have a maximum horizontal socialization level, and thus a minimum vertical socialization. Under polarization, the horizontal socialization level is close to its minimum and thus vertical socialization is close to its maximum. Interestingly, when there is polarization of norms, the larger the majority, the farther away are both norms and socialization levels. Indeed, in order to stick to its preferred behavior (different from the one of the majority) the smaller the minority is, the higher the vertical socialization becomes.

Results change in Section 3.2, where we analyze the case of agents playing always the same game (singular material payoff distribution). Depending on the initial norms, in both complements and substitutes environments, there exist material payoffs such that the society may converge toward norm assimilation or diverge to polarization. Moreover, we show that partial convergence or partial polarization can be sustained in games with strategic complements and substitutes, respectively. In these cases, one community has a norm so strong as to generate a dominant strategy while the other does not have such a strong norm and best replies to the dominant strategy as influenced mostly by material payoffs.

In Section 4, we discuss the model outcomes under general payoff distributions, the role of assortativity on the matching process for the speed of convergence, and possible further development of the model allowing for mixed (complements and substitutes) environments. Section 5 concludes the paper.

1.1 Literature Review

In recent years, a very wide literature about norms and their effect in socio-economic outcomes has emerged. Many works focus on the relationship between norms (or culture) and coordination. Acemoglu and Jackson (2014) study the evolution of a cooperation norm; Dalmazzo et al. (2014) present conditions under which harmful cultural traits can persist in a community; Michaeli and Spiro (2017) address the arising of biased norm when agents, with pressure to conform to each other, play a coordination game; Carvalho (2016) shows how cultural constraints can lead to miscoordination. In Tabellini (2008) agents who are matched together to play a Prisoner Dilemma face a trade-off between individual values (inherited from parents) and material incentives. The main contribution of our work with respect to these papers is that we study the outcome for different classes of games at once and let the norm formation process depend both on the imitation of peers (horizontal socialization) and the transmission of parents (vertical socialization).

The literature about cultural transmission was initiated by Cavalli-Sforza and Feldman (1981) and, in economics, by Bisin and Verdier (2001), where the evolution of cultural traits is the result of parent's socialization choices. Socialization can be vertical (parents), horizontal (peers), and oblique (role models). Along these lines, Bisin and Verdier (2017) study the joint evolution of culture and institutions. In our paper, the socialization is vertical, when parents transmit their preferences to offsprings, and horizontal, when peers interact together to form new norms. In our model the transmitted cultural traits are continuous, as in Panebianco (2014). For a complete theoretical and empirical survey on cultural transmission literature see Bisin and Verdier (2011).

For what concern the effect of norms on the payoff structure, this paper refers to a specific behavioral literature (López-Pérez, 2008; Kessler and Leider, 2012; Kimbrough and Vostroknutov, 2016) where actions depends on the will to adhere to a norm.[6] The main difference is that in our paper agents are affected by a group-specific norm, not necessarily equal for the whole society, so that different players can be subject to different norms.

The concept of cognitive dissonance we use for the dynamics of preferences was introduced in economics by Akerlof and Dickens (1982). Kuran and Sandholm (2008) and Calabuig et al. (2016, 2017, 2018), whose norm formation and norm dynamics are close to ours, have also elements of cognitive dissonance in the updating of norms. In particular, our contribution can be seen as an extension of Kuran and Sandholm (2008) where agents, endowed with their norms, interact in strategic environments, and where the dynamics of norms depends on the interaction between norms and the related equilibrium outcome of games. If we switch-off feedbacks of equilibrium actions on transmitted norms and horizontal socializations, our model boils down to a discrete time version of Kuran and

[6] An alternative viewpoint is that norms imply preferences for a certain distribution of outcomes, i.e. uniform across players.

Sandholm (2008). Having these feedbacks changes the results drastically, for example we are able to reproduce cultural heterogeneity even with fixed communities and complete interaction.

Our model talks also to the literature of identities and oppositional cultures pioneered in economics by Akerlof and Kranton (2000). Kuran and Sandholm (2008) study the tension between cultural integration and multiculturalism, Bisin et al. (2011) focus on the reason that leads to the presence of oppositional cultures, Olcina et al. (2017) address the problem of minorities embedded in a relationships network who decide whether or not to be assimilated to the majority norm. Our main contribution with respect to this literature is to make explicit the effect of different strategic environments.

The literature about indirect evolution (see for example Güth and Yaari, 1992; Güth and Kliemt, 1994; Bester and Güth, 1998; Guttman, 2000) studies environments in which evolutionary selection acts indirectly on preferences. Our approach, even if close in the spirit, is more strongly related to Bisin et al. (2004), where the evolution of preferences is moved by a purely cultural transmission mechanism.

Finally, our model can be applied even in the framework of opinion dynamics (DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2010, 2012) where there is the tension between reaching consensus and disagreement. For example, Yildiz et al. (2013) find in the presence of stubborn agents the reasons of disagreement, Golub and Jackson (2010) study the general conditions for reaching the consensus in a network, and Bolletta and Pin (2019) show how an endogenous network structure can lead to opinion's polarization. In this framework, according to our model, agents form their opinion taking into account both their previous opinion and the one of others. Then, when they are supposed to take decisions, they are affected by both opinions and material rewards. Thus, they update and transmit new opinions taking into account also the experience gained through interaction. The main insight of our work with respect to this literature is that the interplay between material incentives and opinions may be crucial for leading to a consensus or to disagreement.

2 The Model

In this section, we introduce our model for the interplay between norms and strategic interaction. We begin with a general overview.

Consider a society of mass 1 with infinitesimal agents divided into two communities $\mathcal{I} = \{1, 2\}$. Without loss of generality, define $\eta \in \left[\frac{1}{2}, 1\right)$ the size of community 1, the majority. Agents belonging to the same community are assumed to be equal and $i \in \mathcal{I}$ is the representative agent of each community.

Each time period $t \in \mathbb{N} \cup \{0\}$ indexes a generation of agents. We divide a generation into three different sub-periods. In *Stage (y)*, youth, the social coordination game that microfounds the choice of personal norms takes place; in *Stage (a)*, adult age, agents interact by playing games whose payoffs are determined also by personal norms; in *Stage (o)*, old age, norms and socialization levels are transmitted to the next generation.

Stage (y) When young, members of the two communities are endowed with type-specific observable personal norms $\boldsymbol{\theta}_t = (\theta_{1,t}, \theta_{2,t}) \in [0,1]^2$ and socialization levels $\boldsymbol{f}_t = (f_{1,t}, f_{2,t}) \in [0, \bar{f}]^2$, where $\bar{f} \in (0,1)$ is a parameter that represents the maximal level of horizontal socialization. Both characteristics are inherited by the previous generation. Playing a social coordination game, young agents symmetrically choose *ex-post* personal norms $\mathbf{x}_t = (x_{1,t}, x_{2,t}) \in [0,1]^2$.

Stage (a) During their adult age, agents interact in a strategic environment. Agents are randomly matched in pairs to play several symmetric 2×2 games. Different games are available in the same period and each game is played according to a probability distribution γ. Each agent plays with members of both communities, namely a fraction η of times against the majority and a fraction $1 - \eta$ against the minority. Games and population matches are drawn from independent distributions. Norms influence total payoffs and the Nash equilibrium actions emerge as the response of both material payoffs and personal norms $\mathbf{x}_t$. $\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}_t] = (\mathbb{E}_{\eta,\gamma}[A_{1,t}], \mathbb{E}_{\eta,\gamma}[A_{2,t}]) \in [0,1]^2$ is the vector of average equilibrium actions of each community in period t, where $\mathbb{E}_\rho[.]$ is the expectation operator with respect to the measure ρ.

Stage (o) At the end of their life, every agent reproduces asexually giving birth to one child. At this stage, parents transmit new norms and choose how much to socialize their offsprings. During the transmission, parents are assumed to be partially myopic: they are able to anticipate the socialization game of their offspring, *stage (a)*, but they are not able to anticipate their future utility from playing 2×2 games. We model the feedback from the environment (game) to norms as a cultural transmission where the Nash equilibrium action most played in the game contribute to determine the inherited personal norms of the new generation. In particular, cognitive dissonance moves the choice of the norm $\boldsymbol{\theta}_{t+1}$ to be transmitted while cultural substitution moves the choice of the socialization parameter $\boldsymbol{f}_{t+1}$.[7]

In the next section, we start our analysis from the illustration of the norm formation in the young age. Next, we consider the effect of norms on the payoffs of games played in agent's adult life. Finally, we characterize the transmission process. In the first two sections, we avoid the time index to simplify the notation.

[7]Cultural substitution is microfounded in Appendix B.

2.1 Young Age Norms' Formation

In this section, we model agents choice of *ex-post* norms $\mathbf{x}$, stemming from the inherited norms $\boldsymbol{\theta}$ and horizontal socializations $\boldsymbol{f}$. In our model, young agents (children) are active in choosing their own personal norm.

As in many cultural evolution paper (e.g. Kuran and Sandholm, 2008), agent *ex-post* norms $\mathbf{x}$ are the result of social interaction. The general idea is that agents' choice of a norm is affected both by inherited norms and by the average *ex-post* norm chosen by peers, as dependent on the horizontal socializations.[8] Since we have two communities, and we assume that all agents within the same community are equal, inherited norms, socialization levels and, as a result, *ex-post* norms are two-dimensional vectors.[9] The norm x_i is chosen by each agent in community $i \in \mathcal{I}$ by maximizing

$$u_i(\mathbf{x}, \theta_i, f_i) = -\underbrace{f_i(x_i - \mathbb{E}_\eta[x])^2}_{social\ coordination} - \underbrace{(1 - f_i)(x_i - \theta_i)^2}_{group\ (or\ family)\ identity},$$

where $\mathbb{E}_\eta[x]$ is taken over the distribution of individual characteristics and it is the average chosen norm.[10] Since f_i is the horizontal socialization, $(1 - f_i)$ is the vertical socialization. The utility function captures the tension between inherited preferences, θ_i, and coordinating with others: when choosing a personal norm agents want to pick a norm not too different from the one of their peers, depending on their horizontal socialization parameter.[11]

Given the distributions (θ_1, θ_2) and (f_1, f_2), the *ex-post* personal norms (x_1, x_2) are found as the unique symmetric Nash equilibrium of the social interaction game, where agents of the same type choose the same *ex-post* personal norm.

Proposition 1
For all $\boldsymbol{\theta} \in [0,1]^2$ and $\boldsymbol{f} \in [0, \bar{f}]^2$, there exists a unique symmetric Nash Equilibrium $\mathbf{x} \in [0,1]^2$ of the norm formation game with

$$x_i = f_i \left(\mathbb{E}_\eta [\theta] - \frac{cov_\eta[f, \theta]}{(1 - \mathbb{E}_\eta[f])} \right) + (1 - f_i)\theta_i \quad for\ i = 1, 2. \tag{1}$$

[8] This is consistent with sociological literature about social norms, see Bicchieri et al. (2018) for a survey.

[9] Notice that, in principle, agents in the same community can make different choices. However, we focus only on symmetric choices for all the agents of the same community and thus, with an abuse of notation, we use only the community index from the beginning.

[10] $\mathbb{E}_\eta[x]$ can be seen as descriptive norm (Muldoon et al., 2014). Notice that since agents are myopic in their youth they are not able to anticipate future payoffs and thus form their norms taking in consideration only parents and peers pressure and not the subsequent strategic environment.

[11] This formulation is exactly equivalent to the *conformity game* played by children in Vaughan (2013). Moreover, beauty contest like utility function, such as the one used in Morris and Shin (2002), is widely used both in the literature of evolution of cultural traits (Kuran and Sandholm, 2008, among others) as well as in network economics for opinion or belief learning and dynamics (Golub and Jackson, 2012; Bolletta and Pin, 2019; Della Lena, 2019), where it can be seen as a micro-foundation of the so called De Groot model (DeGroot, 1974)

The average norm is

$$\mathbb{E}_\eta[x] = \mathbb{E}_\eta[\theta] - \left(\frac{\mathbb{E}_\eta[f]}{1 - \mathbb{E}_\eta[f]}\right) cov_\eta[f, \theta]. \tag{2}$$

Proof. In the Appendix. □

The result is a generalization of Kuran and Sandholm (2008). In that setting, $\boldsymbol{f}$ is the same for both groups, $cov[f, \theta] = 0$, and thus by (2) $\mathbb{E}_\eta[x] = \mathbb{E}_\eta[\theta]$: the population average *ex-post* norm is equivalent to the population average inherited norm. In our model, the heterogeneity of horizontal socialization introduces a distortion in the distribution of *ex-post* norms. Even a minority, if enough rigid, can make her group norms prevail.

The following corollary expresses equilibrium norms as a convex combination of inherited norms, where p_i is the weight that each community i gives to the inherited norm of community 1 (the majority).

Corollary 1.1
The Nash Equilibrium of Proposition 1 can be written as

$$\begin{cases} x_1 & = p_1\theta_1 + (1 - p_1)\theta_2 \\ x_2 & = p_2\theta_1 + (1 - p_2)\theta_2 \end{cases}, \tag{3}$$

where $p_1 = \frac{(1-f_1)(1-f_2(1-\eta))}{1-f_1\eta-f_2(1-\eta)} \in (0, 1)$, $p_2 = \frac{f_2\eta(1-f_1)}{1-f_1\eta-f_2(1-\eta)} \in (0, 1)$ *and* $p_1 > p_2$ *for all* η, f_1, f_2.

Social interaction makes each agent choose as norm a convex combination between her initial norm and the one of the other community. Weights depend on both types socialization parameters and the majority size η. By taking the difference of p_1 and p_2, it can be easily seen that p_1 is always greater that p_2. Thus if $\theta_1 > \theta_2$, then $x_1 > x_2$ (and viceversa): it is not possible to have a switch of ordering between *ex-ante* and *ex-post* norms. Finally note that $f_i = 0$ implies $x_i = \theta_i$.

2.2 Nash Equilibria for Normal Form Games with Norms

In this section, we model how norms change the payoffs of each one-stage game and study the implication on the game's (pure) Nash equilibria.

In their adult age, agents use norms to make strategic decisions. Their choice is affected both by material and immaterial payoffs. The latter are represented by the willingness to choose an action as indicated by their norms. Some norms are often associated with cooperative environment while others with competitive ones. Below we provide anecdotal examples about norms associated with environments with strategic complementarity or substitutability.

Language (complements) When people interact in a multicultural environment, they have to choose the language to use. On the one hand, there are evident "material" incentives to coordinate on the same language. On the other hand, agents can have different preferences in using a specific language (norms). Preferences for one language can depend on the relative pleasure of using it, on agents ability to speak it, or on other idiosyncrasies.

Attitude toward conflict (substitutes) In competitive interactions for shared resources, the material incentive are to anti-coordinate, the optimal action is to be aggressive, "fight", when the other agent is not, "flight", and viceversa (as in hawks-doves class of games). In this case, the material incentives are as in anti-coordination games.

Work ethics (complements/substitutes) In interacting at the workplace people may face both an environment with strategic complements and substitutes. If we consider a work task that needs a team effort to be accomplished and there is no reward if both agents do not exert a high level of effort, agents have incentives to coordinate and the game is with strategic complements. On the contrary, easy tasks that can be accomplished with the effort of only one agent open the doors for free-riding and the game is with strategic substitutes. These two examples can be thought as, on the one hand, a tough environment where resources are extractable at high labor cost and where agents have to cooperate (complements), and, on the other hand, a flourishing environment where there are abundant and easily extractable resources, in which some agents have the chance to free-ride (substitutes).

We represent the tasks that individuals face with symmetric bi-matrix games where norms interact with material payoffs and lead to the total, material plus immaterial, payoff. For each bi-matrix game Γ, the set of players is $\mathcal{N} = \{r, c\}$ and the action space is defined as $\mathcal{A} = \mathcal{A}_r \times \mathcal{A}_c$ where $\mathcal{A}_r = \mathcal{A}_c = \{1, 0\}$ is the set of actions available for each player (e.g. language A or language B, being aggressive or not, high effort or low effort at the workplace). The material payoff of the bi-matrix game is symmetric and $\pi_r(1, 1) = a$, $\pi_r(0, 1) = b$, $\pi_r(1, 0) = c$, and $\pi_r(0, 0) = d$ (all strictly positive).

We consider norms as preferences over actions, namely, the closer the action to a norm, the higher the perceived utility.[12] Therefore, the experienced ultimate payoffs are, in general, different from the material payoffs. In particular, naming i_r the community of

[12]We could express this ordering of preferences over consequences in the framework of psychological games (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009, among others) where players have belief-dependent motivations (such as intentions-based reciprocity, emotions, or concern with others' opinion); the main difference is that in our framework, the payoff of each agent does not depend on beliefs about others.

player r, for each action profile $\boldsymbol{A} = (A_r, A_c)$ player r payoff function with norms is

$$\Pi_r(\boldsymbol{A}; x_{i_r}) = \begin{cases} x_{i_r}\pi_r(1, A_c) & \text{when} \quad A_r = 1 \\ (1 - x_{i_r})\pi_r(0, A_c) & \text{when} \quad A_r = 0 \end{cases} . \tag{4}$$

The same modification of the material payoff holds also for player c, whose community is denoted as j_c. Table 1 represents the payoff matrix associated with Γ.

		Agent c	
		1	0
Agent r	1	$x_{i_r}a, x_{j_c}a$	$x_{i_r}c, (1 - x_{j_c})b$
	0	$(1 - x_{i_r})b, x_{j_c}c$	$(1 - x_{i_r})d, (1 - x_{j_c})d$

Table 1: Bi-matrix game with norms as preferences over behavior

Concentrating on agent r, when $x_{i_r} = \frac{1}{2}$ the transformation is just a rescaling of payoffs, $\Pi_r\left(\boldsymbol{A}; \frac{1}{2}\right) = \frac{1}{2}\pi_r(\boldsymbol{A})$, and thus has no effect on best replies. Thus, the norm is said to be **neutral**. On the contrary, when $x_{i_r} > 1/2$ the norm is in favor of action 1 so that agent r total payoff for playing action 1 is larger than when the norm is neutral or in favor of action 2, ($x_{i_r} < 1/2$). The larger the norm the larger such an influence. If the norm is **extreme** ($x_{i_r} = 0$ or $x_{i_r} = 1$), the agent always plays the associated action, thus giving no importance to material payoffs.

Modeling the effect of norms as in (4) we capture the idea that norms change preferences over material payoffs, as in Bisin et al. (2004) and Tabellini (2008) but, at the same time, when norms are extreme (0 or 1), the transformation is consistent with the interpretation of Carvalho (2016) where norms restrict agents' strategy set.[13]

Since we are interested in studying the evolution of norms when agents are exposed to different strategic environments, namely when in absence of norms actions are strategic complements or substitutes, we restrict the ordering of the material payoffs a, b, c, d. Measuring with $\bar{b} = \frac{b}{a+b}$ the material force that leads out from the equilibrium $(1, 1)$ and with $\bar{d} = \frac{d}{c+d}$ the material force that pushes toward the equilibrium $(0, 0)$, it is possible to categorize the possible games with material payoffs $\pi(\boldsymbol{A})$ as $\Gamma(\bar{b}, \bar{d})$. In particular

1. Coordination (*strategic complements*): $\bar{b} < \frac{1}{2} < \bar{d}$

[13]An agent that takes into account both material and moral payoffs as in (4) can be seen as "Homo Moralis" in the language of Alger and Weibull (2013). Moreover, the functional form for payoffs (4) is also consistent with one commonly used in the behavioral literature on social norms (López-Pérez, 2008; Kessler and Leider, 2012; Kimbrough and Vostroknutov, 2016) where a cost function c of violating the norm is subtracted to the material payoff: $\Pi_r(\boldsymbol{A}; x_{i_r}) = \pi_r(\boldsymbol{A}) - c(x_{i,r}, \boldsymbol{A}, \pi)$. Indeed, with $c(x_{i_r}, \boldsymbol{A}, \pi) = \pi_r(A_r, A_c)(A_r + x_{i_r}(1 - 2A_r))$ we get exactly equation (4).

2. Anti-Coordination (*strategic substitutes*): $\bar{d} < \frac{1}{2} < \bar{b}$.[14]

Similarly to norms, also $\bar{b}$ and $\bar{d}$ belong to the interval $[0, 1]$. However, they represent material, instead of moral, incentives. If $\bar{b} = \bar{d} = \frac{1}{2}$, then there are no material incentives. A game with both moral and material incentives is denoted as $\Gamma(\bar{b}, \bar{d}, \mathbf{x})$.

We can now proceed with the equilibrium analysis. For simplicity, we assume complete information about material payoffs, norms, and rationality of agents and we use (pure strategy) Nash equilibrium as the solution concept.

The equilibrium analysis relies on the double effect on moral and material incentives. Moral incentives depend on the consistency between norms and actions. Therefore, the final decision depends on the strength of the norm as compared to the two threshold $\bar{b}$ and $\bar{d}$. $\bar{b}$ establishes the minimum strength of the norm for action 1 to be played when the opponent plays 1. $\bar{d}$ establishes the maximum strength of the norm for action 0 to be played when the opponent plays 0.[15]

Define $\hat{A}(\bar{b}, \bar{d}, x) = \hat{A}_r(\bar{b}, \bar{d}, x_{i_r}) \times \hat{A}_c(\bar{b}, \bar{d}, x_{j_c})$ the set of Nash Equilibria with norms.

Proposition 2 *Given the game with norms $\Gamma(\bar{b}, \bar{d}, x)$:*

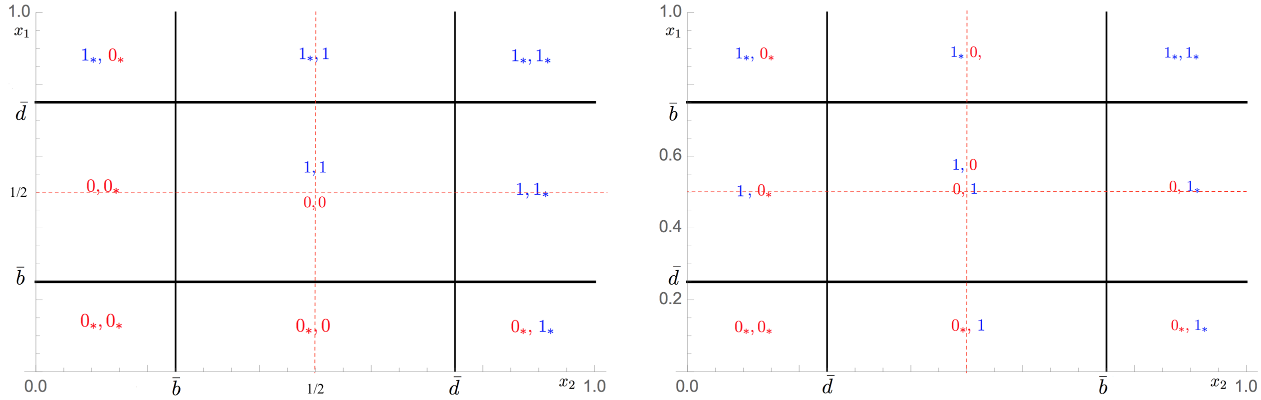
- *If $x_{i_r} > \bar{b}$ and $x_{j_c} > \bar{b}$, then $(1, 1) \in \hat{A}(\bar{b}, \bar{d}, x)$.*
- *If $x_{i_r} < \bar{d}$ and $x_{j_c} < \bar{d}$, then $(0, 0) \in \hat{A}(\bar{b}, \bar{d}, x)$.*
- *If $x_{i_r} > \bar{d}$ and $x_{j_c} < \bar{b}$, then $(1, 0) \in \hat{A}(\bar{b}, \bar{d}, x)$.*
- *if $x_{i_r} < \bar{b}$ and $x_{j_c} > \bar{d}$, then $(0, 1) \in \hat{A}(\bar{b}, \bar{d}, x)$.*

Proof. In the Appendix. □

The results of Proposition 2 are represented in Figures 1 where equilibrium actions played by agents belonging to the two different communities are shown as a function of norms and for different strategic environments: complements (left) or substitutes (right). The set of Nash Equilibria depends on the position of threshold values $\bar{d}$ and $\bar{b}$. Figure 1(a) represents a game with strategic complements, $\bar{b} < \frac{1}{2} < \bar{d}$. Figure 1(b) represents a game with strategic substitutes, $\bar{d} < \frac{1}{2} < \bar{b}$. On the main diagonal there are Nash Equilibria when agents have the same norm. When the action is marked with the subscript $_*$, it is dominant. As expected, for x_{i_r} and x_{j_c} in the neighborhood of neutral norms ($x_{i_r} = x_{j_c} = \frac{1}{2}$), the games have the same equilibria as the corresponding game without norms, while as x_{i_r} and x_{j_c} move away from $\frac{1}{2}$ the games have have different equilibria.

[14] Notice that restricting the ordering of $\bar{b}, \bar{d}, \frac{1}{2}$ can be used to characterize even Prisoner Dilemma and Efficient Dominant Strategy Equilibrium games, where $\frac{1}{2} < \min\{\bar{b}, \bar{d}\}$ and $\frac{1}{2} > \max\{\bar{b}, \bar{d}\}$, respectively. Our general analysis applies also to these other games but we focus on complements vs substitutes. See also the discussion in Section 4.

[15] This is consistent, even if in a totally different framework, with Eshel et al. (1998), who found that the imitation dynamics depends only upon the values α and β which are strictly related respectively with our $\bar{b}$ and $\bar{d}$.



(a) (b)

Figure 1: Nash Equilibria as a function of norms (x_1, x_2). The row player (vertical axis) belongs to community 1, the column player (horizontal axis) belongs to community 2. (a) Complements, $\bar{b} < \frac{1}{2} < \bar{d}$ and (b) Substitutes, $\bar{d} < \frac{1}{2} < \bar{b}$. A star denotes that the action is dominant.

When $x_{i_r} > \max\{\bar{d}, \bar{b}\}$ or $x_{i_r} < \min\{\bar{d}, \bar{b}\}$, r has a dominant strategy, respectively playing 1 or playing 0. Otherwise, if $\min\{\bar{d}, \bar{b}\} < x_{i_r} < \max\{\bar{d}, \bar{b}\}$, then r has not a dominant strategy and reacts to the action of c. Thus if both x_{i_r} and x_{j_c} are between $\min\{\bar{d}, \bar{b}\}$ and $\max\{\bar{d}, \bar{b}\}$, then we have multiple equilibria. We now interpret the results with respect the two anecdotal examples of language and attitude toward conflict.

Language (complements) The interaction of norms with the incentive for coordination may result in different Nash equilibria. If the norm is strong and equal in both communities, norms and incentives are aligned on one language. If one norm is strong, say of community 1, and the other is mild, of community 2, community 1 uses always the preferred language while community 2 uses the language preferred by community 1 when matched with 1 and can use either language when matched with a player of community 2. If norms of the two communities are strong but different, the game could instead become an anti-coordination one, where each member of a community uses only its most preferred language (polarization). The stronger the material incentives to coordinates the stronger should be the norm to obtain polarization.

Attitude toward conflict (substitutes) Norms in the form of attitude toward conflict could change the Nash equilibria as well. If, for example the norms of players are strong and aligned, the game could result in one where both players flight (or fight). On the contrary, two strong and not aligned norms have the effect of selecting one equilibrium with anti-coordination. Finally, if the norm of one community is mild while the one of other community is strong then the members of that community anti-coordinate with members of the other one, and are indifferent in interacting among themselves.

2.3 Cultural Transmission

At the end of each time period t, given the action played during their adult age, agents transmit new norms $\boldsymbol{\theta}_{t+1}$ to their offsprings and decide how much let them socialize with the peers by transmitting a horizontal socialization $\boldsymbol{f}_{t+1}$. In this section, we model both transmissions as a function of generation t norms, $\mathbf{x}_t$, socializations, $\boldsymbol{f}_t$, and average actions chosen when playing the one-stage games.

The average action depends on the distribution of games each agent is playing in her adult age. We shall assume that in each period agents play different games. In particular, we name γ a probability distribution on the space of vectors $(\bar{b}, \bar{d})$ and assume that the game with payoffs $\Gamma(\bar{b}, \bar{d}; \mathbf{x}_t)$ is played with probability $\gamma(\bar{b}, \bar{d})$. In order to study the effect of different strategic environment on norms dynamics, we further assume that games played belong always to the same environment. In complements environments $\bar{b} \in \left[0, \frac{1}{2}\right]$ and $\bar{d} \in \left[\frac{1}{2}, 1\right]$, viceversa for substitutes environments $\bar{b} \in \left[\frac{1}{2}, 1\right]$ and $\bar{d} \in \left[0, \frac{1}{2}\right]$.

We focus our analysis on the two extreme cases: (i) γ is a uniform distribution on sets of $(\bar{b}, \bar{d})$ with support $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$ for complements and $\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$ for substitutes. (ii) γ is a point distribution on an element of the set $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$ for complements and $\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$ for substitutes.

2.3.1 Formation of Transmitted Norms

First, we assume that agents try to reduce the cognitive dissonance that arises if there is no consistency between their original preference over actions $\mathbf{x}_t$ and their average behavior. The latter, $\mathbb{E}_{\eta,\gamma}[A_{i,t}]$, is computed taking into account that agents are randomly matched to play against the whole population and that they face one-stage games drawn from the distribution γ. Thus, an agent belonging to the majority, community $i = 1$, will be matched η time with his own type and $1 - \eta$ with others, viceversa for a player belonging to the minority. Moreover, each agent in her adult life plays an infinite number of games with different payoffs $(\bar{b}, \bar{d})$ distributed according to γ.

We model cognitive dissonance by letting the *ex ante* personal norms of each community i generation $t + 1$, $\theta_{i,t+1}$, to move towards the transmitted actions $\mathbb{E}_{\eta,\gamma}[A_{i,t}]$ as dependent on the influence of games on norms $\lambda \in (0, 1)$:

$$\theta_{i,t+1} = (1 - \lambda)x_{i,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{i,t}] =: \zeta_i(x_{i,t}, \mathbb{E}_{\eta,\gamma}[A_{i,t}]). \tag{5}$$

If $\lambda = 0$ the strategic environment has no effect on the evolution of norms, similarly the case considered by Kuran and Sandholm (2008). Otherwise, the inherited norm of the next generation depends directly on the norm of the parent $x_{i,t}$ and indirectly also on the strategic environment through the average equilibrium actions in $\mathbb{E}_{\eta,\gamma}[A_{i,t}]$. Parents want to transmit to their offsprings norms that are a combination of their norms and of what they have learned to be the best action in the specific strategic environment they face. Note that an equivalent motivation for this form of the feedback is a sociological story: since children, in the vertical socialization process, are able to observe both norms and behavior, they inherit norms similar to those of their parents but biased toward parental behavior.

In order to find the average action played by the representative agent of one community, $\mathbb{E}_{\eta,\gamma}[A_{i,t}]$, we have to integrate actions with respect to both population and payoffs measure, η and γ.[16] In particular, all the time agents play a game with a unique Nash equilibrium the transmitted action is uniquely defined to be the Nash equilibrium one. When, instead, the played game has not a unique Nash equilibrium, we assume indifference on the type of action that is transmitted. Thus, each time an agent of community i plays against an agent of community j in period t, the transmitted action is (assuming without loss of generality that she is the row player)

$$A_{i,j,t} = \begin{cases} \hat{A}_{r,t}(\bar{b}, \bar{d}, x) & \text{if the Nash equilibrium is unique} \\ \frac{1}{2} & \text{otherwise} \end{cases} .$$

Notice that, whenever there are multiple equilibria, the feedback from actions to norms of a player of community i when playing with a player of community j in date t is $A_{i,j,t} = \frac{1}{2}$.[17]

In Figure 2, we show average actions for agents of both communities for different (x_1, x_2), in both complements and substitutes game. Integrating over games and communities, we get the vector of average actions:[18]

$$\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}_t] = (\mathbb{E}_{\eta,\gamma}[A_{1,t}], \mathbb{E}_{\eta,\gamma}[A_{2,t}]) =: \boldsymbol{\varphi}(\mathbf{x}_t). \tag{6}$$

2.3.2 Formation of Horizontal Socialization

The dynamics of the socialization level depends on the outcome of the strategic interaction and it is modeled assuming cultural substitution.[19]

[16]Notice that, due to the assumption of independence the order of integration does not affect the result and $\mathbb{E}_{\eta,\gamma}[.] = \mathbb{E}_{\gamma,\eta}[.]$

[17]A different possibility could be to consider the Mixed Nash equilibrium. However, in several empirical works the approximation of $\frac{1}{2}$ is widely used in presence of multiple equilibria (Bjorn and Vuong, 1984; Kooreman, 1994; Soetevent and Kooreman, 2007).

[18]It is possible to take into account different levels of assortativity in the matching, see discussion in Section 4.

[19]Under cultural substitution "parents have fewer incentives to socialize their children the more widely dominant are their values in the population" Bisin and Verdier (2001).

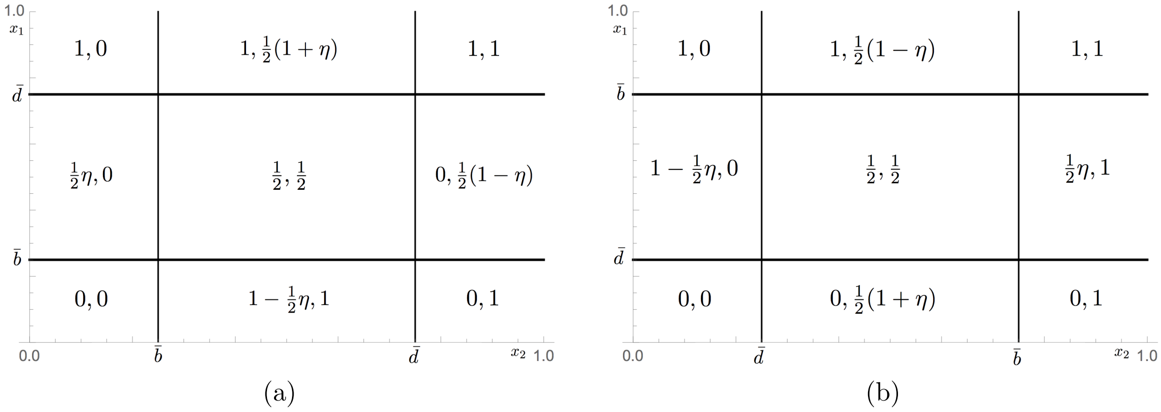


Figure 2: $\mathbb{E}_{\eta}[A_{1,t}], \mathbb{E}_{\eta}[A_{2,t}]$ in strategic settings with (a) Strategic complements $\left(\bar{b} \leq \frac{1}{2} \leq \bar{d}\right)$, (b) Strategic substitutes $\left(\bar{d} \leq \frac{1}{2} \leq \bar{b}\right)$. As in Figure 1, we assume that the row player (vertical axis) belongs to community 1 while the column player (horizontal axis) belongs to community 2.

We assume that the more their action is close to the average of the society, the more agents let their offsprings horizontally socialize with the peers (the less they vertically socialize them). Given the average action in the whole society,

$$\bar{A}_t = \eta \mathbb{E}_{\eta,\gamma}[A_{1,t}] + (1-\eta)\mathbb{E}_{\eta,\gamma}[A_{2,t}],$$

the transmitted horizontal socialization for a representative agent of community i is

$$f_{i,t+1} = \bar{f}(1 - |\mathbb{E}_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|) =: \psi_i(\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}_t]). \tag{7}$$

The horizontal socialization of community i depends directly on the differences between the average action of the agent and the average action of the whole society. $\bar{f}$ is the maximum possible flexibility parameter of the society, namely the highest level of horizontal socialization of a community toward the whole society, and thus also to the other community. When the distance between actions is maximal, the components of $\boldsymbol{f}$ go to their lower bound. When the distance is minimal (agents play all the same action) both components of $\boldsymbol{f}$ reach the upper bound: $\boldsymbol{f} = (\bar{f}, \bar{f})$, where $\bar{f} < 1$.

It is important to underline that we have assumed cultural substitution and the specific functional form (7) to have tractability in the dynamics. However, in Appendix B, we microfound it showing that (7) is consistent with partially myopic parents, with utility function as in Panebianco (2014), who are able to anticipate the socialization game of their offspring, but they are not able to anticipate their utility when adult.

3 Norm and Socialization Level Dynamics

In this section, we analyze the joint dynamics of norms, Nash equilibrium actions, and socialization levels for environments where, under material payoffs, actions are strategic complements or substitutes.

As anticipated in Section 2.1, the chosen norm $\mathbf{x}$ depends directly on inherited norm $\boldsymbol{\theta}$ and horizontal socialization $\boldsymbol{f}$. Introducing the time dimension and using eq. (1) for a representative agent of both communities $i \in \mathcal{I}$, we define the vector function $\boldsymbol{\upsilon}(.)$ such that

$$\mathbf{x}_{t+1} = \boldsymbol{\upsilon}(\boldsymbol{\theta}_{t+1}, \boldsymbol{f}_{t+1}). \tag{8}$$

Combining equations (6 - 8) we get the dynamics of our model

$$\mathbf{x}_t \xrightarrow{\varphi(.)} \mathbb{E}_{\eta,\gamma}[\boldsymbol{A}_t] \xrightarrow{\boldsymbol{\zeta}(.),\ \boldsymbol{\psi}(.)} (\boldsymbol{\theta}_{t+1}, \boldsymbol{f}_{t+1}) \xrightarrow{\boldsymbol{\upsilon}(.)} \mathbf{x}_{t+1}.$$

or

$$\mathbf{x}_{t+1} = \boldsymbol{\upsilon}(\boldsymbol{\zeta}(\mathbf{x}_t, \boldsymbol{\varphi}(\mathbf{x}_t)), \boldsymbol{\psi}(\boldsymbol{\varphi}(\mathbf{x}_t))) =: \boldsymbol{\Xi}(\mathbf{x}_t). \tag{9}$$

We define $\mathcal{E}$ the set of steady states of (9), $\mathcal{E} := \{\mathbf{x}^* \in [0,1]^2 : \mathbf{x} = \boldsymbol{\Xi}(\mathbf{x})\}$. From each steady-state norm $\mathbf{x}^*$, we can derive the corresponding steady-state actions $(\mathbb{E}_{\eta,\gamma}[A_1^*], \mathbb{E}_{\eta,\gamma}[A_2^*])$, the steady-state horizontal socialization levels (f_1^*, f_2^*), and the steady-state equilibrium weights (p_1^*, p_2^*) that link *ex-post* and *ex-ante* norms, as defined in Corollary 1.1. First, we provide a relation between norms and average actions at the steady state.

Proposition 3

Given the norm dynamics in (9), for all strategic environments, each steady-state norm $\mathbf{x}^*$ *solves*

$$\begin{cases} x_1^* = \phi_1^* \mathbb{E}_{\eta,\gamma}[A_1^*] + (1-\phi_1^*)\mathbb{E}_{\eta,\gamma}[A_2^*] \\ x_2^* = \phi_2^* \mathbb{E}_{\eta,\gamma}[A_1^*] + (1-\phi_2^*)\mathbb{E}_{\eta,\gamma}[A_2^*] \end{cases}, \tag{10}$$

where $\phi_1^* = \frac{p_1^* - (p_1^* - p_2^*)(1-\lambda)}{1-(p_1^*-p_2^*)(1-\lambda)}$ *and* $\phi_2^* = \frac{p_2^*}{1-(p_1^*-p_2^*)(1-\lambda)}$.

Proof. In the Appendix. □

In a steady state, norms are a convex combination of the average actions played by the agents of the two communities. Weights (ϕ_1^*, ϕ_2^*) depend on the steady state norms $\mathbf{x}^*$ through horizontal socialization levels $\boldsymbol{f}^*$ and weight p_1^* and p_2^*. The size of the majority η and the influence of games on norms λ also play a role.

Next, we find steady-state norms, and characterize their stability, for both complements and substitutes environments. Below, we consider uniform distributions of material payoffs. In Section 3.2, we consider point distributions.

3.1 Uniform Distribution of Material Payoffs

In this section the payoff distribution γ is uniform distribution on sets of material payoffs $(\bar{b}, \bar{d})$. We consider separately cases where material payoffs imply that actions are strategic complements or substitutes.

3.1.1 Strategic Complements

In their adult age, agents face an environment with strategic complementarity when each game they play has material payoffs with $0 \leq \bar{b} \leq \frac{1}{2} \leq \bar{d} \leq 1$. Assuming that payoffs are uniformly distributed, we choose $\bar{b}$ and $\bar{d}$ uniformly in $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, respectively, and independently for each other.

As we show below, there exists (only) steady states where both communities share the same, extreme or neutral, norm. Here, the level of horizontal socialization is maximal and communities play the same Nash equilibrium action. The set of steady states norms is $\mathcal{E}^d := \{(1,1), (0,0), \left(\frac{1}{2}, \frac{1}{2}\right)\}$, all with maximal socialization $\bar{f}$, and with average actions $\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}^*]$ equal to $(1,1)$, $(0,0)$, and $\left(\frac{1}{2}, \frac{1}{2}\right)$, respectively. Moreover the steady state where both communities have the same strong norm, $(1,1)$ and $(0,0)$, are stable, while the state with neutral norms, $\left(\frac{1}{2}, \frac{1}{2}\right)$, is a saddle.

Proposition 4 (Complements)
For all $\eta \in [\frac{1}{2}, 1)$, $\bar{f} \in (0,1)$, and $\lambda \in (0,1)$, if $(\bar{b}, \bar{d})$ is uniformly distributed in the set $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$, then the set of steady states is $\mathcal{E}^d$. Moreover, $(1,1)$ is asymptotically stable with basin of attraction at least $\left(\frac{1}{2}, 1\right]^2$, $(0,0)$ is asymptotically stable with the basin of attraction at least $\left[0, \frac{1}{2}\right)^2$, and $\left(\frac{1}{2}, \frac{1}{2}\right)$ is a saddle.

Proof. in the Appendix $\square$

In an environment with strategic complementarity, if payoffs are uniformly distributed over $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$, the two stable long-run outcomes are assimilation toward one extreme norm, either 0 or 1. In each stable equilibrium, the prevailing norm solves the Nash equilibrium selection issue typical of coordination games. In our framework, whether norms converge to 1 or 0 depends on the initial norms and thus, in turn, on basins of attraction. For all initial conditions, but those on the saddle path leading to the neutral norm $\frac{1}{2}$, both community converge to the same norm, as well as to the corresponding Nash equilibrium action, and the horizontal socialization is at its maximum.

3.1.2 Strategic Substitutes

We now discuss results for an environment with strategic substitutability where agents play several games in which $0 \leq \bar{d} \leq \frac{1}{2} \leq \bar{b} \leq 1$. Whereas in an environment with strategic complements both games played by agents in the youth age (the norm formation game) and the adult age (the 2×2 strategic interaction) favor coordination among agents, with strategic substitutes two opposite forces are at play: coordination in youth and anti-coordination in the adult age. As we shall see, the relative strength of the two, as dependent on the relationship between the parameter that describes the maximum possible horizontal socialization ($\bar{f}$) and the strength of the cognitive dissonance (λ), determines long-run outcomes.

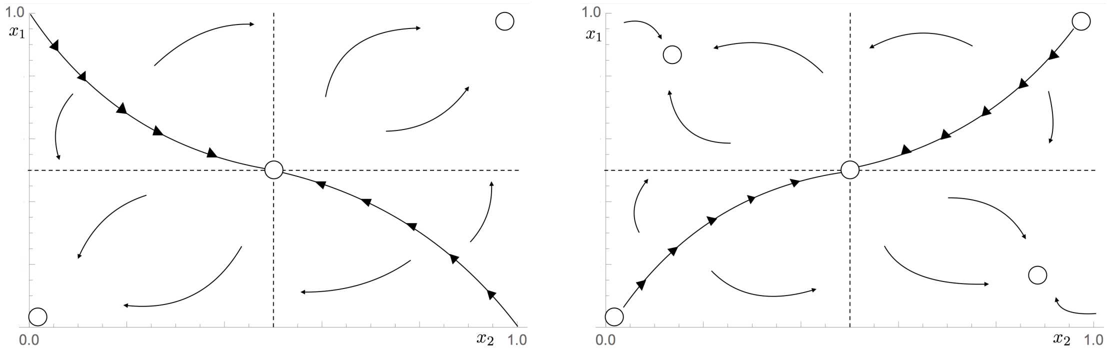
In games with substitutes, other than the set of steady states where both communities share the same norm and the socialization level is at its maximum, $\mathcal{E}^d$, there exists a set of polarized steady states where the two communities use different norms and the horizontal socialization level is low, $\mathcal{E}^p := \{(\overset{\diamond}{x}_1, \overset{\diamond}{x}_2) \in \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right], (\overset{\circ}{x}_1, \overset{\circ}{x}_2) \in \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]\}$. To maintain analytic tractability, in the next proposition we provide results only when the two communities have the same size, $\eta = \frac{1}{2}$.

Proposition 5 (Substitutes)
For $\eta = \frac{1}{2}$ and $\lambda \in (0,1)$, if $(\bar{b}, \bar{d})$ is uniformly distributed in the set $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$, then there exist $\hat{f} \geq \frac{\lambda}{2+\lambda}$ such that

- *If $\bar{f} > \hat{f}$, then the set of steady states is $\mathcal{E}^d$ and $\left(\frac{1}{2}, \frac{1}{2}\right)$ is the globally stable steady state.*
- *If $\bar{f} < \hat{f}$, then the set of steady states is $\mathcal{E}^d \cup \mathcal{E}^p$ and the strong norm states $(0,0)$ and $(1,1)$ are unstable. Moreover:*
  - *If $0 < \bar{f} < \frac{\lambda}{2+\lambda}$, then $\left(\frac{1}{2}, \frac{1}{2}\right)$ is a saddle, $(\overset{\diamond}{x}_1, \overset{\diamond}{x}_2)$ is asymptotically stable with basin of attraction at least $\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$, $(\overset{\circ}{x}_1, \overset{\circ}{x}_2)$ is asymptotically stable with basin of attraction of at least $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$;*
  - *If $\frac{\lambda}{2+\lambda} < \bar{f} < \hat{f}$, then all norms in $\mathcal{E}^p \cup \left(\frac{1}{2}, \frac{1}{2}\right)$ are asymptotically stable.*

Proof. In the Appendix □

In an environment with substitutes and uniformly distributed material payoffs, there still exist steady states with *assimilation* as in the previous (complements) case, however they are unstable. Stable steady states depend on the tension between the young age social interaction and the strategic environment faced in the adult age. If the young age social interaction is very strong also across the two communities ($\bar{f} > \hat{f}$), the only stable long run outcome is the erosion of norms $\left(\frac{1}{2}, \frac{1}{2}\right)$ (as in Calabuig et al., 2016). On the contrary, with low social interaction, $\bar{f} < \hat{f}$, there exist stable steady states with polarization. In the latter, norms are polarized, e.g. $\overset{\diamond}{x}_1 > \frac{1}{2}$ and $\overset{\diamond}{x}_2 < \frac{1}{2}$, agents of the two communities play on average opposite actions $(\mathbb{E}_{\eta,\gamma}[\overset{\diamond}{A}_1], \mathbb{E}_{\eta,\gamma}[\overset{\diamond}{A}_2]) \in \left[\overset{\diamond}{x}_1, 1\right] \times \left[0, \overset{\diamond}{x}_2\right]$, and the more the actions are polarized the more the horizontal socialization is low. (symmetrically for $\overset{\circ}{x}_1$ and $\overset{\circ}{x}_2$).



(a) (b)

Figure 3: Norm dynamics for (a) Strategic Complements and (b) Strategic Substitutes when $\bar{f} < \frac{\lambda}{2+\lambda}$.

When $\bar{f} < \hat{f}$, whether the polarized steady states are the only possible long-run outcomes or norm neutrality is also a possible long-run outcome depends on the relative size of the horizontal socialization response to behavioral heterogeneity ($\bar{f}$) and agents' cognitive dissonance (λ). The higher the cognitive dissonance, the higher the effect of strategic substitutes in the adult age games, making norm neutrality a saddle. This is always the case when $\bar{f} = 0$, so that horizontal socialization does not play a role and children care only about inherited norms, leading to fully polarized norms, either $(\overset{\diamond}{x}_1 = 1, \overset{\diamond}{x}_2 = 0)$ or $(\overset{\circ}{x}_1 = 0, \overset{\circ}{x}_2 = 1)$.

The main implication of Propositions 4 and 5 is that different strategic environments lead to different long-run norms and equilibrium behavior. An environment with strategic complementarity always leads to norms that are strong and homogeneous across both communities. Instead, in environments with strategic substitutability, there exist stable steady states with polarization of norms or, when the norm is homogeneous it is neutral in both communities. Whether polarization or erosion of norms is the long-run outcomes depends on the relative size of cognitive dissonance and horizontal socialization.

3.2 Point Distribution of Material Payoffs

In the previous section, we have shown that with strategic complements and uniformly distributed payoffs stable norms are strong and homogeneous. An empirical observation is that an environment that favors coordination is not always enough to ensure the occurrence of complete assimilation of the minority or to avoid the occurrence of polarization in a society. The process of cultural integration may fail to achieve complete assimilation, having the resilience of cultural traits, or even lead to norms' polarization.[20] Bisin and

[20]We refer to Berry (1997) and Ryder et al. (2000) (among others) for the terminology about cultural *assimilation*, *integration*, *marginalization*, and *polarization* (also *separation*). They proposed a concept of minority's self-identification, based on a two-dimensional framework, which takes into account for differences in both adaptation and interaction processes between the minority and the dominant culture.

Verdier (2011) offer a review of empirical examples of cultural heterogeneity and resilience of cultural traits: the slow rate of immigrants' integration in Europe and US, the persistence of 'ethnic capital' in second- and third-generation immigrants, e.g. minorities' strongly attachment to their original languages and cultural traits.

In this section, we show how to reconcile our model of norm formation with these empirical observations. We do so by considering limit cases of point distributions of material payoffs, namely by imposing γ to be singular on the set $(\bar{b}, \bar{d})$. We shall show that polarization can occur even in environments with complements while mild cultural heterogeneity can occur both in environments with complements, favoring convergence of norms, and in environments with substitutes, favoring divergence of norms.

As with the uniform distribution of payoffs the exact nature of these stable norms depends on whether the strategic interaction taking place in the adult age is characterized by strategic substitutes or complements. We start our analysis with the latter.[21]

3.2.1 Strategic Complements

Figure 4 shows all the possible steady states in environments with complements. The first result is that two stable steady states with polarization (and minimal horizontal socialization) can exist depending on the value of $\bar{b}$ and $\bar{d}$ and on the norms' distance between the two communities. For simplicity, we now formally present only one of the two steady states with polarization (the upper left blue dot in Figure 4), the other (the lower right blue dot) can be easily derived by symmetry.

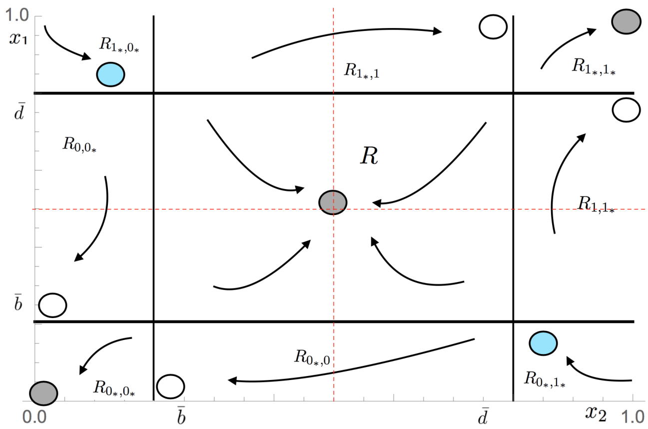


Figure 4: Steady States of $\mathbf{x}$ for complements with $\bar{b} = 0.2$, $\bar{d} = 0.8$, $\bar{f} = 0.3$, $\eta = 0.5$, $\lambda = 0.6$.

The existence of steady states depend on the relative size of material incentives $(\bar{b}, \bar{d})$ and average actions weights defined in Proposition 3. If $\bar{b} > \phi_1^*$ and $\bar{d} < \phi_2^*$ and initial norms belong to the region $R_{1_*,0_*}$, then the dynamics described in (9) converges to the steady state (ϕ_1^*, ϕ_2^*), with socialization $(\bar{f}\eta, \bar{f}(1-\eta))$. Norms are polarized, horizontal

[21] We keep the analysis at a descriptive level, more details can be found in the Appendix.

socialization is at its minimum, and agents belonging to different communities always play different actions, $\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}^*] = (1, 0)$. This occurs despite in all rounds agents are playing a coordination game, according to material payoffs.

The result is driven from the fact that if games played in adult life have always the same non degenerate payoffs, then there exist initial norms, $x_{1,0}$ high enough and $x_{2,0}$ low enough, that sustain the equilibrium with anti-coordination. Playing this equilibrium leads, by cognitive dissonance, to norms polarization.

If agents interaction features complementarity another class of steady states can exist (white in Figure 4), where there is cultural *integration* but not complete *assimilation*. In such steady states, only the agents of one community have a well-defined group-specific norm, which induce them to always play a specific action as dominant strategy. Agents belonging to the other community have a neutral norm. When the latter are matched among themselves, they face the original coordination problem, while they conform to the behavior of agents with a strong norm whenever they encounter them. This is a clear example of *integration*, namely there is convergence toward a homogeneous norm, but, at the same time, the identity is not totally lost as in *assimilation*.

Real life examples of this result are linguistic choices between immigrants and natives. Natives always use their own language. Agents belonging to a linguistic minority start with a different norm and, after a long interaction with natives, they end up using the two languages indifferently, but conforming with the natives whenever they interact with them.

In steady states with *integration*, we have two sources of symmetry. One is with respect to the community with a well defined norm, the other with respect to the action played. Therefore there can exist up to four steady states of this type. Again we formally characterize only one of these equilibria. Let us focus on the region $R_{1_*,1}$, were there can exist a steady state where $\bar{b} < x_2^* < \bar{d}$ and $\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}] = \left(1, \frac{1}{2}(1+\eta)\right)$. Notice that with point distribution steady state with *intergation* but not *assimilation* are stable, provided they exist. However, there are values of material incentives $(\bar{b}, \bar{d})$ such that they do not exist, as shown in Figure 5. On the contrary, steady states with *assimilation* always exist (and are always stable).[22]

The difference between *assimilation* and *integration* equilibria is important with respect to different policy goals. For example, sometimes a policy is considered successful only when minorities (immigrants) completely lose their previous norms or culture and are assimilated; instead in other circumstances the resilience of cultural traits can be considered socially desirable, in these cases the policymaker reaches its goal if the minority integrates with the majority but keeps some of their cultural traits. In this second case, there is partial convergence and there is still room for a multicultural society.

[22] In Appendix A.8 we prove that with point distributions of payoffs all steady states are asymptotically stable. The issue is thus whether a steady state exists or not, given material payoffs, rather than if it is stable.

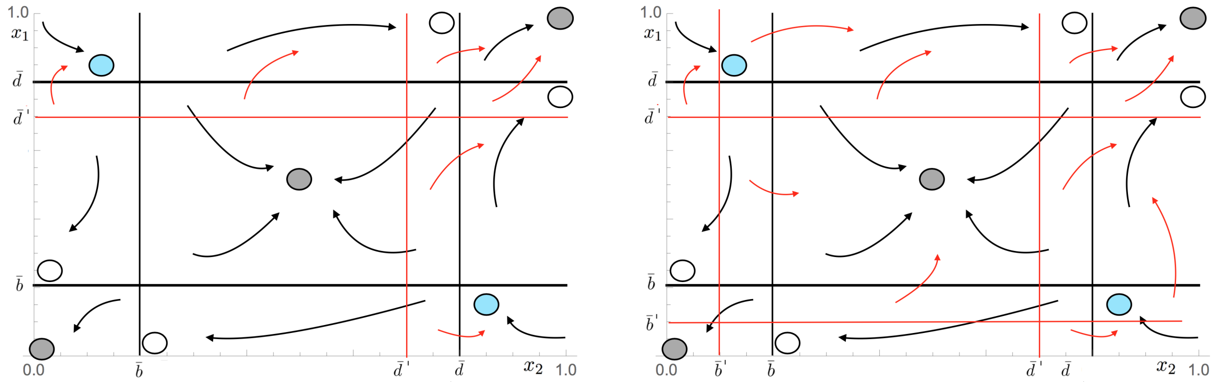


Figure 5: Steady states of the norm dynamics for strategic complements with ($\bar{b} = 0.2, \bar{d} = 0.8$) and ($\bar{d}' = 0.7, \bar{b}' = 0.1$). Left panel: change from d to d'. Right panel: change from d to d' and from b to b'. In both plots, $\bar{f} = 0.3$, $\eta = 0.5$, $\lambda = 0.6$.

Moreover, from a policy point of view, it is interesting to appraise the long-run effects due to a change in game incentives (Figure 5). We observe that moving the material incentives $\bar{b}$ and $\bar{d}$ can significantly affect the social outcome. In Figure 5 (left panel) we can see how diminishing $\bar{d}$ to $\bar{d}'$ the basin of attraction of the steady state with assimilation (grey) becomes much wider and integration (white) disappears. Figure 5 (right panel) shows that in such a case, moving the two material incentives together, $\bar{d}$ to $\bar{d}'$ and $\bar{b}$ to $\bar{b}'$, it is possible to reach assimilation even if the communities start off having initial norms that are polarized. This sheds further light on the relationship between uniform and point distribution. In fact, if $(\bar{b}, \bar{d})$ is not a single point, but moves in the whole space $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$, then the steady states presented in this section are not robust to game change.

3.2.2 Strategic Substitutes

Figure 6 shows all the possible steady states in environments with strategic substitutes when γ is singular. The main difference with the case of uniform distribution is that steady states with assimilation can be stable. Comparing the long-run norms with the case of point distribution and strategic complements, steady states with cultural *integration* do not exist. Moreover, in environments with substitutes, steady states may exist (white in Figure 6) where one community has a well defined norm while the members of the other community have a norm that, when they are matched among themselves, does not induce preferences over actions. Namely, interacting in their own community agents are indifferent on the action to play, but when matched with the other community they act in the opposite way. In a sense, we can talk about *marginalization*, in fact, there is a partial polarization of norms with low socialization across communities.

In environments with strategic substitutes policies that aim to reach cultural integration are not possible. Cultural assimilation is still possible but it is not reachable starting from polarized norms, even if we weaken material incentives to anti-coordinate. These incentives determine if there exists steady states with marginalization or polarization. Lowering material incentives enlarges the region of initial norms that lead to norm neutrality. Differently from the uniform distribution of material payoffs (Proposition 5) the relative size of the maximum horizontal socialization $\bar{f}$ and cognitive dissonance parameter λ do not play a role for stability but only for the speed of convergence.[23]

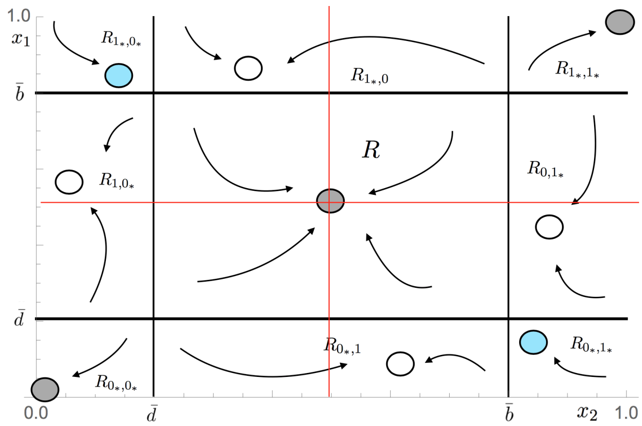


Figure 6: Steady states of the norm dynamics for strategic substitutes with $\bar{b} = 0.8$, $\bar{d} = 0.2$, $\bar{f} = \eta = 0.5$, $\lambda = 0.75$. In the extreme case where $\bar{b} = \bar{d} = 0.5$, the neutral norm $(0.5, 0.5)$ is the only steady state and it is globally stable.

[23] See the Proof of Proposition A.8 in the Appendix for details.

4 Discussion

In this section we discuss possible extensions of the model.

Payoffs Distribution This paper is a first attempt to analyze the effect of different strategic environment on the dynamics of norms. While, we study only extreme cases for the payoffs' distribution (uniform and point), our analysis can be generalized to many others payoffs distributions. If the distribution is still uniform, important parameters to understand the existence and stability of stable norms are the boundary points (highest and lowest) of the payoffs distribution of material incentives $(\bar{b}, \bar{d})$. In the interior points of the support the dynamics is the same as studied in 3.1, while outside results discussed in 3.2 hold. Considering probability distributions different from the uniform one, the analysis is less straightforward as it depends on the density in the tails. Our conjecture is that it is possible to establish threshold values over which the density is vanishing and such that the dynamics is the same as with a uniform distribution defined within these threshold. Thresholds should depend, non trivially, on the whole distribution, including moments higher than the second. Although all these possible intermediate cases could be interesting to study, their analysis is beyond the scope of this paper.

Assortativity We have studied the case of perfect random matching without taking into account the possibility of having assortative matching. In order to consider different levels of assortativity, it is enough to consider a parameter ϵ that can assume values less than η for the minority ($1 - \eta$ for the majority) and add it to the probability of being matched with agents belonging to the same community. This generalization does not affect results.[24] In environments with complements, the only effect of a higher level of assortativity is to slow down the convergence to the steady state norm for minority and speed it up for the majority, thus the assimilation of norms occurs at a slower pace. On the contrary, in environments with substitutes, a higher level of assortativity decreases the speed of convergence to the steady state, and thus to the polarization. This would suggest us that, from the prospective of a policy maker who wants to favor assimilation, in environments with complements is better to facilitate across-communities interactions, while in environments with substitutes is better to avoid them.

Mixed Environments In this paper we keep the class of strategic environment fixed to complements or substitutes. An extension of the model is to allow changes in the class of strategic environment. Our conjecture is that norms may not converge and generate cycles. The same results should be found also when the strategic environment resembles Prisoner Dilemma. In fact, according to our preliminary analysis, depending on the

[24]The only difference is to replace η with $\eta' = \eta + \epsilon$ for the majority and with $\eta'' = \eta - \epsilon$ for the minority. Since the only condition relevant for the proofs is $\eta < 1$, the change does not affect the dynamics.

material payoffs parametrization, the Prisoner Dilemma with norms becomes either a game with strategic complements or a game with strategic substitutes.

This extension is particularly relevant to study more complex environments where some the same norm is applied to both cooperative and competitive settings. Such an extension could provide a foundation of the fact that if a society faces many tasks that require a joint effort, then its agents develop more cooperative norms. This is also in line with the recent empirical evidence suggesting that institutions, by changing material payoffs, can lead to a crowding out of norms (Lowes et al., 2017).

5 Conclusion

In this paper, we study a cultural transmission model where the relationship between norms and strategic environments is made explicit. Agents divided into two communities form their community norm by taking into account both the norm received by their previous generation and the average norm of the society. The relative strength of the two forces is regulated by a horizontal socialization parameter. The norm received by the previous generation depends on the average equilibrium actions played in the game under the hypothesis of minimization of cognitive dissonance. We derive conditions under which norm assimilation is reachable or not. Provided games material payoffs are randomly distributed but preserve their strategic setting (complements/substitutes), the norm dynamics converges to assimilation in environments with strategic complements and to polarization or neutrality in environments with strategic substitutes. Moreover, when specific material payoffs are chosen, provided initial conditions show enough heterogeneity, we are able to obtain the rise of oppositional cultures and situations of cultural heterogeneity. For example, we show that even if the material payoffs provides incentives to coordinate, it is still possible to obtain integration but not assimilation or, even, polarization. At the same time, in environments where material payoffs provide incentive to anti-coordinate, it is still possible to reach assimilation or only partial polarization.

References

Aarts, H. and Dijksterhuis, A. (2003). The silence of the library: environment, situational norm, and social behavior. *Journal of Personality and Social Psychology*, 84(1):18.

Acemoglu, D. and Jackson, M. O. (2014). History, expectations, and leadership in the evolution of social norms. *The Review of Economic Studies*, 82(2):423–456.

Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, 90(4):599–617.

Akerlof, G. A. and Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *American Economic Review*, 72(3):307–319.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Alger, I. and Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological bulletin*, 91(1):3.

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology*, 46(1):5–34.

Bester, H. and Güth, W. (1998). Is altruism evolutionarily stable? *Journal of Economic Behavior & Organization*, 34(2):193–209.

Bicchieri, C., Muldoon, R., and Sontuoso, A. (2018). Social norms. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Ethnic identity and labour market outcomes of immigrants in europe. *Economic Policy*, 26(65):57–92.

Bisin, A., Topa, G., and Verdier, T. (2004). Cooperation as a transmitted cultural trait. *Rationality and Society*, 16(4):477–507.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Bisin, A. and Verdier, T. (2011). The economics of cultural transmission and socialization. In *Jess Benhabib, Alberto Bisin, and Matt Jackson (eds) Handbook of social economics.*, volume 1, chapter 9. Elsevier.

Bisin, A. and Verdier, T. (2017). On the joint evolution of culture and institutions. *NBER Working Paper Series.*

Bjorn, P. A. and Vuong, Q. H. (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. *SSWP 537, California Institute of Technology.*

Bolletta, U. and Pin, P. (2019). Polarization when people choose their peers. *mimeo.*

Calabuig, V., Olcina, G., and Panebianco, F. (2016). The erosion of personal norms and cognitive dissonance. *Applied Economics Letters*, 23(18):1265–1268.

Calabuig, V., Olcina, G., and Panebianco, F. (2017). The dynamics of personal norms and the determinants of cultural homogeneity. *Rationality and Society*, 29(3):322–354.

Calabuig, V., Olcina, G., and Panebianco, F. (2018). Culture and team production. *Journal of Economic Behavior & Organization*, 149:32–45.

Carvalho, J.-P. (2016). Coordination and culture. *Economic Theory*, 64(3):1–27.

Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach.* Number 16. Princeton University Press.

Cooper, J. and Fazio, R. H. (1984). A new look at dissonance. *Advances in experimental social psychology*, 17:229–268.

Dalmazzo, A., Pin, P., and Scalise, D. (2014). Communities and social inefficiency with heterogeneous groups. *Journal of Economic Dynamics and Control*, 48:410–427.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

Della Lena, S. (2019). Non-bayesian social learning and the spread of misinformation in networks. *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series No*, 9.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968.

Eshel, I., Samuelson, L., and Shaked, A. (1998). Altruists, egoists, and hooligans in a local interaction model. *American Economic Review*, 88(1):157–179.

Festinger, L. (1962). *A theory of cognitive dissonance*. Stanford university press.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and economic Behavior*, 1(1):60–79.

Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.

Güth, W. and Kliemt, H. (1994). Competition or co-operation: on the evolutionary economics of trust, exploitation and moral attitudes. *Metroeconomica*, 45(2):155–187.

Güth, W. and Yaari, M. (1992). An evolutionary approach to explain reciprocal behavior in a simple strategic game. In *Explaining Process and Change–Approaches to Evolutionary Economics*. U. Witt. Ann Arbor: University Michigan Press.

Guttman, J. M. (2000). On the evolutionary stability of preferences for reciprocity. *European Journal of Political Economy*, 16(1):31–50.

Hayek, F. A. (1967). *Notes on the Evolution of Systems of Rules of Conduct*. University of Chicago Press Chicago.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies*, 59(1):63–80.

Kessler, J. B. and Leider, S. (2012). Norms and contracting. *Management Science*, 58(1):62–77.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kooreman, P. (1994). Estimation of econometric models of some discrete games. *Journal of Applied Econometrics*, 9(3):255–268.

Kuran, T. and Sandholm, W. H. (2008). Cultural integration and its discontents. *The Review of Economic Studies*, 75(1):201–228.

López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior*, 64(1):237–267.

Lowes, S., Nunn, N., Robinson, J. A., and Weigel, J. L. (2017). The evolution of culture and institutions: Evidence from the kuba kingdom. *Econometrica*, 85(4):1065–1091.

Michaeli, M. and Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9(1):152–216.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Muldoon, R., Lisciandra, C., Bicchieri, C., Hartmann, S., and Sprenger, J. (2014). On the emergence of descriptive norms. *Politics, philosophy & economics*, 13(1):3–22.

Olcina, G., Panebianco, F., and Zenou, Y. (2017). Conformism, social norms and the dynamics of assimilation. *IZA Discussion Papers, No. 11436.*

Panebianco, F. (2014). Socialization networks and the transmission of interethnic attitudes. *Journal of Economic Theory*, 150:583–610.

Ryder, A. G., Alden, L. E., and Paulhus, D. L. (2000). Is acculturation unidimensional or bidimensional? a head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of personality and social psychology*, 79(1):49.

Schwartz, S. H. (1977). Normative influences on altruism. *Advances in experimental social psychology*, 10:221–279.

Soetevent, A. R. and Kooreman, P. (2007). A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics*, 22(3):599–624.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950.

Vaughan, D. (2013). Nurture vs. nurture: Endogenous parental and peer effects and the transmission of culture. *Working Papers, Banco de México.*

Yildiz, E., Ozdaglar, A., Acemoglu, D., Saberi, A., and Scaglione, A. (2013). Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation (TEAC)*, 1(4):19.

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

Young, H. P. (1998). Social norms and economic welfare. *European Economic Review*, 42(3):821–830.

A Appendix

A.1 Proof of Proposition 1

In order to simplify the notation, we consider payoffs for a representative agent of each community. The payoff of a generic agent belonging to community $i \in \mathcal{I}$ is

$$u_i(\mathbf{x}, \theta_i, f_i) = -f_i \left(x_i - \underbrace{(\eta x_1 + (1-\eta) x_2)}_{\mathbb{E}_\eta[x]} \right)^2 - (1 - f_i)(x_i - \theta_i)^2.$$

Since each agent is negligible in the population, she does not affect the whole average. Thus, the first order condition $\frac{\partial u_i(\mathbf{x}, \theta_i)}{\partial x_i} = 0$ gives

$$\begin{aligned} -2f_i \left(x_i - (\eta x_1 + (1-\eta)x_2)\right) - 2(1-f_i)(x_i - \theta_i) &= 0 \\ f_i \left(x_i - (\eta x_1 + (1-\eta)x_2)\right) + (1-f_i)(x_i - \theta_i) &= 0, \end{aligned}$$

leading to

$$x_i = f_i \mathbb{E}_\eta[x] + (1 - f_i)\theta_i. \tag{11}$$

Taking expectations of (11) on both sides we get

$$\begin{aligned} \mathbb{E}_\eta[x] &= \mathbb{E}_\eta[f \cdot \mathbb{E}_\eta[x] + (1-f)\theta] \\ \mathbb{E}_\eta[x] &= \mathbb{E}_\eta[f \cdot \mathbb{E}_\eta[x]] + \mathbb{E}_\eta[\theta] - \mathbb{E}_\eta[f \cdot \theta] \\ \mathbb{E}_\eta[x] - \mathbb{E}_\eta[f] \cdot \mathbb{E}_\eta[x] &= \mathbb{E}_\eta[\theta] - E_\eta[f \cdot \theta] \\ \mathbb{E}_\eta[x] - \mathbb{E}_\eta[f] \cdot \mathbb{E}_\eta[x] &= \mathbb{E}_\eta[\theta] - \mathbb{E}_\eta[f] \cdot \mathbb{E}_\eta[\theta] - cov_\eta[f, \theta] \\ (1 - \mathbb{E}_\eta[f]) \cdot \mathbb{E}_\eta[x] &= (1 - \mathbb{E}_\eta[f]) \cdot \mathbb{E}_\eta[\theta] - cov[f, \theta] \\ \mathbb{E}_\eta[x] &= \mathbb{E}_\eta[\theta] - \frac{cov_\eta[f, \theta]}{(1 - \mathbb{E}_\eta[f])}. \end{aligned}$$

Substituting $\mathbb{E}_\eta[x]$ in (11) we find the optimal action of each player belonging to community i as a function of the distributions of θ and f

$$x_i = f_i \left(E_\eta[\theta] - \frac{cov_\eta[f, \theta]}{(1 - E_\eta[f])} \right) + (1 - f_i)\theta_i. \tag{12}$$

□

A.2 Proof of Corollary 1.1

From the ex-post norm formation rule (12) we obtain for agent $i = 1$

$$x_1 = f_1 \left(\eta\theta_1 + (1-\eta)\theta_2 - \frac{cov_\eta[f,\theta]}{1 - f_1\eta - f_2(1-\eta)} \right) + (1-f_1)\theta_1.$$

Computing the covariance leads to

$$\begin{aligned} cov_\eta[f,\theta] &= \mathbb{E}_\eta\left[(f - \mathbb{E}_\eta[f])(\theta - \mathbb{E}_\eta[\theta])\right] \\ &= \eta(1-\eta)^2(f_1 - f_2)(\theta_1 - \theta_2) + \eta^2(1-\eta)(f_2 - f_1)(\theta_2 - \theta_1) \\ &= \eta(1-\eta)(f_1 - f_2)(\theta_1 - \theta_2) \end{aligned}$$

so that

$$\begin{aligned} x_1 &= f_1 \left(\eta\theta_1 + (1-\eta)\theta_2 - \frac{\eta(1-\eta)(f_1-f_2)(\theta_1-\theta_2)}{1 - f_1\eta - f_2(1-\eta)} \right) + (1-f_1)\theta_1 \\ &= \underbrace{\frac{f_1(1-\eta)(1-f_2)}{1 - f_1\eta - f_2(1-\eta)}}_{1-p_1} \theta_2 + \underbrace{\frac{(1-f_1)(1-f_2(1-\eta))}{1 - f_1\eta - f_2(1-\eta)}}_{p_1} \theta_1. \end{aligned}$$

The same can be computed for agent $i = 2$.

□

A.3 Proof of Proposition 2

We solve the generic game with norms as described in Table 1. The best-replies are

$$\hat{A}_r(A_c = 1; \bar{b}, \bar{d}, x_{i_r}) = \begin{cases} 1 \ if \ x_{i_r} > \bar{b} \\ 0 \ if \ x_{i_r} < \bar{b} \end{cases} \quad \text{and} \quad \hat{A}_r(A_c = 0; \bar{b}, \bar{d}, x_{i_r}) = \begin{cases} 1 \ if \ x_{i_r} > \bar{d} \\ 0 \ if \ x_{i_r} < \bar{d} \end{cases} .$$

Since the game is symmetric, these are also the best reply of agent c. Looking for the fixed-point of the best replies we find the Nash Equilibria.

□

A.4 Corollary 2.1

Before we proceed with other proofs, we provide a partition of the norm space $[0, 1]^2$ where Nash equilibria are as described in Figure 1.

Corollary 2.1 The regions of norms x_{i_r} and x_{j_c} in which different Nash equilibria emerge are the following:

$$\begin{cases} R_{1_*,1_*} = \{(x_{i_r}, x_{j_c}) : x_{i_r} > max\{\bar{d}, \bar{b}\} \wedge x_{j_c} > max\{\bar{d}, \bar{b}\}\} \\ R_{0_*,0_*} = \{(x_{i_r}, x_{j_c}) : x_{i_r} < min\{\bar{d}, \bar{b}\} \wedge x_{j_c} < min\{\bar{d}, \bar{b}\}\} \\ R = \{(x_{i_r}, x_{j_c}) : min\{\bar{d}, \bar{b}\} < x_{i_r} < max\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\}\} \\ R_{1_*,0_*} = \{(x_{i_r}, x_{j_c}) : x_{i_r} > max\{\bar{d}, \bar{b}\} \wedge x_{j_c} < min\{\bar{d}, \bar{b}\}\} \\ R_{0_*,1_*} = \{(x_{i_r}, x_{j_c}) : x_{i_r} < min\{\bar{d}, \bar{b}\} \wedge x_{j_c} > max\{\bar{d}, \bar{b}\}\} \\ R_{1_*,1} = \{(x_{i_r}, x_{j_c}) : x_{i_r} > max\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\} \wedge \bar{d} > \bar{b}\} \\ R_{1_*,0} = \{(x_{i_r}, x_{j_c}) : x_{i_r} > max\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\} \wedge \bar{b} > \bar{d}\} \\ R_{1,1_*} = \{(x_{i_r}, x_{j_c}) : x_{j_c} > max\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{i_r} < max\{\bar{d}, \bar{b}\} \wedge \bar{d} > \bar{b}\} \\ R_{0,1_*} = \{(x_{i_r}, x_{j_c}) : x_{j_c} > max\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{i_r} < max\{\bar{d}, \bar{b}\} \wedge \bar{b} > \bar{d}\} \\ R_{0_*,0} = \{(x_{i_r}, x_{j_c}) : x_{i_r} < min\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\} \wedge \bar{d} > \bar{b}\} \\ R_{0_*,1} = \{(x_{i_r}, x_{j_c}) : x_{i_r} < min\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\} \wedge \bar{b} > \bar{d}\} \\ R_{0,0_*} = \{(x_{i_r}, x_{j_c}) : x_{i_r} < min\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{j_c} < max\{\bar{d}, \bar{b}\} \wedge \bar{d} > \bar{b}\} \\ R_{1,0_*} = \{(x_{i_r}, x_{j_c}) : x_{j_c} < min\{\bar{d}, \bar{b}\} \wedge min\{\bar{d}, \bar{b}\} < x_{i_r} < max\{\bar{d}, \bar{b}\} \wedge \bar{b} > \bar{d}\} \end{cases}$$

A.5 Proof of Proposition 3

Substituting the dynamics of $\boldsymbol{\theta}$ in equation (3) we get

$$\begin{cases} x_{1,t+1} = p_{1,t+1}[(1-\lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + (1-p_{1,t+1})[(1-\lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]] \\ x_{2,t+1} = p_{2,t+1}[(1-\lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + (1-p_{2,t+1})[(1-\lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]] \end{cases},$$

where $p_{1,t} = \frac{(1-f_{1,t})(1-f_{2,t}(1-\eta))}{1-f_{1,t}\eta-f_{2,t}(1-\eta)}$, $p_{2,t} = \frac{f_{2,t}\eta(1-f_{1,t})}{1-f_{1,t}\eta-f_{2,t}(1-\eta)}$.

Substituting and computing the steady state we obtain

$$\begin{cases} x_1^* = \phi_1^*\mathbb{E}_{\eta,\gamma}[A_1^*] + (1-\phi_1^*)\mathbb{E}_{\eta,\gamma}[A_2^*] \\ x_2^* = \phi_2^*\mathbb{E}_{\eta,\gamma}[A_1^*] + (1-\phi_2^*)\mathbb{E}_{\eta,\gamma}[A_2^*] \end{cases}, \tag{13}$$

where $\phi_1^* = \frac{p_1^*-(p_1^*-p_2^*)(1-\lambda)}{1-(p_1^*-p_2^*)(1-\lambda)}$ and $\phi_2^* = \frac{p_2^*}{1-(p_1^*-p_2^*)(1-\lambda)}$ or, in terms of (f_1^*, f_2^*),

$$\begin{cases} \phi_1^* = \frac{(1-f_1^*)f_2\eta+\lambda(1-f_1^*-f_2^*+f_1^*f_2^*)}{f_1(1-\eta)+f_2^*\eta-f_1^*f_2^*+\lambda(1-f_1^*-f_2^*+f_1^*f_2^*)} \\ \phi_2^* = \frac{f_2^*\eta(1-f_1)}{f_1(1-\eta)+f_2^*\eta-f_1^*f_2^*+\lambda(1-f_1^*-f_2^*+f_1^*f_2^*)} \end{cases}.$$

□

A.6 Proof of Proposition 4

To verify that $(1,1)$, $(0,0)$, and $\left(\frac{1}{2},\frac{1}{2}\right)$ are steady states it is enough to substitute them in (13). Due to the dynamics for horizontal socialization, namely

$$\begin{cases} f_1 = \bar{f}(1-(1-\eta)|(\mathbb{E}_{\eta,\gamma}[A_{1,t}] - \mathbb{E}_{\eta,\gamma}[A_{2,t}])|) \\ f_2 = \bar{f}(1-\eta|(\mathbb{E}_{\eta,\gamma}[A_{2,t}] - \mathbb{E}_{\eta,\gamma}[A_{1,t}]])|) \end{cases}, \tag{14}$$

when norms, and thus average actions, are strong or neutral, $\mathbb{E}_{\eta,\gamma}[A_{1,t}] = \mathbb{E}_{\eta,\gamma}[A_{2,t}]$ leading to $f_1^* = f_2^* = \bar{f}$.

In order to study the dynamics and show that no other steady state is present, we need to explicit $\mathbb{E}_{\eta,\gamma}[A_t]$ as a function of x_t. Let's first focus on the region $\{(x_{1,t}, x_{2,t}) : x_{1,t} > \frac{1}{2}, x_{2,t} > \frac{1}{2}\}$.

Assume $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (\frac{1}{2}, 1)$, then

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = \gamma(\bar{d} < x_{1,t}) + \gamma(\bar{d} > x_{1,t} \ \wedge \ \bar{d} < x_{2,t}) \left(1 - \frac{1}{2}\eta\right) + \gamma(\bar{d} > x_{1,t} \ \wedge \ \bar{d} > x_{2,t}) \frac{1}{2}$$

Since payoffs are uniformly distributed in the region characterized by strategic complementarity

- $\gamma(\bar{d} < x_1) = 2\left(x_1 - \frac{1}{2}\right)$
- $\gamma(\bar{d} > x_1 \wedge \bar{d} < x_2) = 4\left(1 - x_1\right)\left(x_2 - \frac{1}{2}\right)$
- $\gamma(\bar{d} > x_1 \wedge \bar{d} > x_2) = 4\left(1 - x_1\right)\left(1 - x_2\right)$

so that

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = 2\left(x_{1,t} - \frac{1}{2}\right) + 4\left(1 - x_{1,t}\right)\left(x_{2,t} - \frac{1}{2}\right)\left(1 - \frac{1}{2}\eta\right) + 4\left(1 - x_{1,t}\right)\left(1 - x_{2,t}\right)\frac{1}{2}.$$

$$= -1 + 2x_{1,t} + 2x_{2,t} - 2x_{1,t}x_{2,t} + (1 - x_{1,t} - 2x_{2,t} + 2x_{1,t}x_{2,t})\eta$$

A similar computation can be performed for $\mathbb{E}_{\eta,\gamma}[A_{2,t}]$.

Next, we shall verify that if $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (\frac{1}{2}, 1)$, then $\min\{x_{1,t}, x_{2,t}\} < \min\{x_{1,t+1}, x_{2,t+1}\} \leq 1$ for all $t \in \mathbb{N}$. This is enough to prove that $(1,1)$ is the only steady state in the region $\{(x_1, x_2) : x_1 \in (\frac{1}{2}, 1), x_2 \in (\frac{1}{2}, 1)\}$ and that it is asymptotically stable with basin of attraction at least as big as the region itself.

We first prove that $\mathbb{E}_{\eta,\gamma}[A_{1,t}] > x_{1,t}$ and $\mathbb{E}_{\eta,\gamma}[A_{2,t}] > x_{2,t}$.

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_{1,t}] > x_{1,t}$:

$$\begin{aligned}
-1 + 2x_{1,t} + 2x_{2,t} - 2x_{1,t}x_{2,t} + (1 - x_{1,t} - 2x_{2,t} + 2x_{1,t}x_{2,t})\eta &> x_{1,t} \\
-1 + x_{1,t} + 2x_{2,t} - 2x_{1,t}x_{2,t} + (1 - x_{1,t} - 2x_{2,t} + 2x_{1,t}x_{2,t})\eta &> 0 \\
\underbrace{(x_{1,t} - 1)}_{>0}\underbrace{(1 - 2x_{2,t})}_{>0}\underbrace{(1 - \eta)}_{>0} &> 0, \qquad \textit{always satisfied.}
\end{aligned}$$

By symmetry (the two groups have different sizes but it is enough to replace η with $(1-\eta)$ for the minority), we can repeat the reasoning for $x_{2,t}$ and show that $\mathbb{E}_{\eta,\gamma}[A_{2,t}] > x_{2,t}$. In particular, the expression for the average action is

$$\mathbb{E}_{\eta,\gamma}[A_{2,t}] = x_{2,t} + (2x_{1,t} - 1)(1 - x_{2,t})\eta\,.$$

By eq. (3) both $x_{1,t}$ and $x_{2,t}$ are a convex combination of $\theta_{1,t}$ and $\theta_{2,t}$. Moreover, if $\mathbb{E}_{\eta,\gamma}[A_{1,t}] > x_{1,t}$ and $\mathbb{E}_{\eta,\gamma}[A_{2,t}] > x_{2,t}$, then by eq. (5) both $\theta_{1,t} > x_{1,t}$ and $\theta_{2,t} > x_{2,t}$. Assume w.l.o.g. $x_{1,t} \geq x_{2,t}$. Then, eqs. (3) and (5) imply $x_{1,t+1} > x_{2,t}$ and $x_{2,t+1} > x_{2,t}$, independently on the weights $(p_{1,t}, p_{2,t})$ of the convex combination in (3). More in general, $\min\{x_{1,t+1}, x_{2,t+1}\} > \min\{x_{1,t}, x_{2,t}\}$ for all $t \in \mathbb{N}$. Moreover, by eqs. (3) and (5) $\min\{x_{1,t+1}, x_{2,t+1}\} \leq 1$. Thus we can conclude that there are not steady states in the region $\{(x_1, x_2) : x_1 \in (\frac{1}{2}, 1), x_2 \in (\frac{1}{2}, 1)\}$ and that $\lim_{t\to\infty} \min\{x_{1,t}, x_{2,t}\} = 1$. The latter together with the fact that norms are bounded above by 1 implies that $(1, 1)$ is asymptotically stable. The proof of the asymptotic stability of $(0, 0)$, as well as the fact that there are no steady states in $\{(x_{1,t}, x_{2,t}) : x_{1,t} \in \left(0, \frac{1}{2}\right), x_{2,t} \in \left(0, \frac{1}{2}\right)\}$, proceeds along the same lines.

We now turn to the region of initial conditions $\{(x_{1,t}, x_{2,t}) : x_{1,t} \in \left(\frac{1}{2}, 1\right), x_{2,t} \in \left(0, \frac{1}{2}\right)\}$ to show that it does not contain steady states, so that the dynamics either converges to $\left(\frac{1}{2}, \frac{1}{2}\right)$ or moves in another region.

Assume $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (0, \frac{1}{2})$, then

$$E_{\eta,\gamma}[A_{1,t}] = \gamma(\bar{d} < x_{1,t}) + \gamma(\bar{d} > x_{1,t} \;\wedge\; \bar{b} > x_{2,t})\frac{1}{2}\eta + \gamma(\bar{d} > x_{1,t} \wedge \bar{b} < x_{2,t})\frac{1}{2}.$$

Since payoffs are uniformly distributed in the region characterized by strategic complementarity

- $\gamma(\bar{d} < x_1) = 2\left(x_1 - \frac{1}{2}\right)$
- $\gamma(\bar{d} > x_1 \;\wedge\; \bar{b} > x_2) = 2(1 - x_1)(1 - 2x_2)$
- $\gamma(\bar{d} > x_1 \wedge \bar{b} < x_2) = 2(1 - x_1)2x_2$

so that

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = 2\left(x_{1,t} - \frac{1}{2}\right) + 2(1-x_{1,t})(1-2x_{2,t})\frac{1}{2}\eta + 2(1-x_{1,t})2x_{2,t}\frac{1}{2}$$
$$= 2x_{1,t} - 1 + \eta(1-x_{1,t})(1-2x_{2,t}) + 2(1-x_{1,t})x_{2,t}.$$

Similarly, we can compute

$$\mathbb{E}_{\eta,\gamma}[A_{2,t}] = \gamma(\bar{d} < x_{1,t} \ \wedge \ \bar{b} < x_{2,t})\frac{1}{2}(1+\eta) + (\bar{d} > x_{1,t} \wedge \bar{b} < x_{2,t})\frac{1}{2}\gamma$$
$$= 2(1+\eta)\left(x_{1,t} - \frac{1}{2}\right)x_{2,t} + 2(1-x_{1,t})x_{2,t}.$$

We shall show below that for all norms $\{(x_{1,t}, x_{2,t}) : x_1 \in \left(\frac{1}{2}, 1\right), x_2 \in \left(0, \frac{1}{2}\right)\}$ $x_{2,t} < \mathbb{E}_{\eta,\gamma}[A_{1,t}] < x_{1,t}$ and $x_{2,t} < \mathbb{E}_{\eta,\gamma}[A_{2,t}] < x_{1,t}$. Using equation (3), the former implies $x_{2,t} < \theta_{1,t+1} < x_{1,t}$ while the latter implies $x_{2,t} < \theta_{2,t+1} < x_{1,t}$. Equation (5) then implies $x_{1,t+1} < x_{1,t}$ and $x_{2,t+1} > x_{2,t}$, proving the result. (in what follows we remove the time index to simplify the notation)

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_1] > x_1$:

$$2x_1 - 1 + \eta(1-x_1)(1-2x_2) + 2(1-x_1)x_2 > x_1$$
$$x_1 - 1 + \eta(1-x_1)(1-2x_2) + 2(1-x_1)x_2 > 0$$
$$(1-x_1)(-1+\eta(1-2x_2)+2x_2) > 0$$

$$\underbrace{(1-x_1)}_{>0}\underbrace{(1-\eta)}_{>0}\underbrace{(2x_2-1)}_{<0} < 0, \qquad \textit{always satisfied.}$$

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_2] > x_2$:

$$2(1+\eta)\left(x_1 - \frac{1}{2}\right)x_2 + 2(1-x_1)x_2 > x_2$$
$$\left(2(1+\eta)\left(x_1 - \frac{1}{2}\right) + 2(1-x_1)\right)x_2 > x_2$$
$$(+2\eta x_1 - \eta + 1)\,x_2 > x_2$$

$$\underbrace{(\eta\underbrace{(2x_1-1)}_{>0}+1)}_{>1}\,x_2 > x_2, \qquad \textit{always satisfied.}$$

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_1] > x_2$

$$\begin{aligned} 2x_1 - 1 + \eta(1-x_1)(1-2x_2) + 2(1-x_1)x_2 &> x_2 \\ 2x_1 - 1 + \eta(1-x_1)(1-2x_2) + 2(1-x_1)x_2 - x_2 &> 0 \\ 2x_1 - 1 + (1-x_1)(\eta(1-2x_2) + 2x_2) - x_2 &> 0 \end{aligned}$$

$$2x_1 - 1 - x_2 + (1-x_1)(\eta + 2x_2(1-\eta)) > 0, \quad \textit{always satisfied.}$$

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_2] < x_1$:

$$\begin{aligned} 2(1+\eta)\left(x_1 - \frac{1}{2}\right)x_2 + 2(1-x_1)x_2 &< x_1 \\ 2x_1x_2 - x_2 + 2\eta x_1x_2 - \eta x_2 + 2x_2 - 2x_1x_2 &< x_1 \\ x_2 + 2\eta x_1x_2 - \eta x_2 &< x_1 \\ x_2(1-\eta+2\eta x_1) &< x_1 \\ x_2 - x_2\eta + 2\eta x_1x_2 - x_1 &< 0 \\ (x_2 - x_1) + x_2\eta(2\eta x_1 - 1) &< 0. \end{aligned}$$

The left-hand side is maximized by $\eta = 1$. Thus, we should verify that

$$\begin{aligned} (x_2 - x_1) + x_2(2x_1 - 1) &< 0 \\ 2x_1x_2 - x_1 &< 0 \\ x_1 \underbrace{(2x_2 - 1)}_{<0} &< 0, \quad \textit{always satisfied.} \end{aligned}$$

The proof that there are no steady states in $\{(x_{1,t}, x_{2,t}) : x_{1,t} \in \left(0, \frac{1}{2}\right), x_{2,t} \in \left(\frac{1}{2}, 1\right)\}$ proceeds along the same lines.

A.7 Proof of Proposition 5

To verify that $(1,1)$, $(0,0)$, and $\left(\frac{1}{2}, \frac{1}{2}\right)$ are steady states one can proceed as in the proof of Proposition 4. Before showing the existence of other fixed points, and assessing their stability, we prove that $(1,1)$ and $(0,0)$ are unstable.

Let us consider the region $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (\frac{1}{2}, 1)$. Assume $x_{1,t} > (\frac{1}{2}, 1)$ and $x_{2,t} > (\frac{1}{2}, 1)$, then

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = \gamma(\bar{b} < x_{1,t}) + \gamma(\bar{b} > x_{1,t} \wedge \bar{b} < x_{2,t})\frac{1}{2}\eta + \gamma(\bar{b} > x_{1,t} \wedge \bar{b} > x_{2,t})\frac{1}{2}.$$

Since payoffs are uniformly distributed in the region characterized by strategic substitutability

- $\gamma(\bar{b} < x_1) = 2x_1 - 1$
- $\gamma(\bar{b} > x_1 \wedge \bar{d} < x_2) = 4(1 - x_1)(x_2 - \frac{1}{2})$
- $\gamma(\bar{b} > x_1 \wedge \bar{b} > x_2) = 4(1 - x_1)(1 - x_2)$

so that

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = (2x_{1,t} - 1) + 4(1 - x_{1,t})(x_{2,t} - \frac{1}{2})\frac{1}{2}\eta + 4(1 - x_{1,t})(1 - x_{2,t})\frac{1}{2}.$$

By eq. (3) both $x_{1,t}$ and $x_{2,t}$ are convex combinations of $\theta_{1,t}$ and $\theta_{2,t}$. The latter depends directly on $\mathbb{E}_{\eta,\gamma}[A_{1,t}]$ and $\mathbb{E}_{\eta,\gamma}[A_{2,t}]$, respectively, as shown in eq. (5). In particular, $\mathbb{E}_{\eta,\gamma}[A_1] < x_1$ and $\mathbb{E}_{\eta,\gamma}[A_2] < x_2$ are sufficient conditions to ensure that $\min\{x_{1,t+1}, x_{2,t+1}\} < \min\{x_{1,t}, x_{2,t}\}$. By symmetry the opposite holds when $x_{1,t} \in (0, \frac{1}{2})$ and $x_{2,t} \in (0, \frac{1}{2})$. Thus, $(1, 1)$ and $(0, 0)$ are unstable.

Let us verify that $\mathbb{E}_{\eta,\gamma}[A_1] < x_1$:

$$2x_1 - 1 + 2\eta(1 - x_1)\left(x_2 - \frac{1}{2}\right) + 2(1 - x_1)(1 - x_2) < x_1$$

$$x_1 - 1 + 2\eta(1 - x_1)\left(x_2 - \frac{1}{2}\right) + 2(1 - x_1)(1 - x_2) < 0$$

$$\underbrace{(1 - x_1)}_{>0}\underbrace{(1 - 2x_2)}_{<0}\underbrace{(1 - \eta)}_{>0} < 0, \quad \textit{always.}$$

By symmetry $\mathbb{E}_{\eta,\gamma}[A_2] < x_2$ and, as we have argued above, both $x_{1,t}$ and $x_{2,t}$ decrease.

We now turn to the existence of other fixed points and to their stability, as well as to the stability of $\left(\frac{1}{2}, \frac{1}{2}\right)$, by analyzing the norms dynamics in the region where $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (0, \frac{1}{2})$ and, symmetrically, $x_{1,t} \in (0, \frac{1}{2})$ and $x_{2,t} \in (\frac{1}{2}, 1)$.

Recall that the norm dynamics is

$$\begin{cases} x_{1,t+1} = p_{1,t+1}[(1 - \lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + (1 - p_{1,t+1})[(1 - \lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]] \\ x_{2,t+1} = p_{2,t+1}[(1 - \lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + (1 - p_{2,t+1})[(1 - \lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]] \end{cases},$$

where $p_{1,t} = \frac{(1-f_{1,t})(1-f_{2,t}(1-\eta))}{1-f_{1,t}\eta-f_{2,t}(1-\eta)}$, $p_{2,t} = \frac{f_{2,t}\eta(1-f_{1,t})}{1-f_{1,t}\eta-f_{2,t}(1-\eta)}$, and from (7)

$$\begin{cases} f_{1,t+1} = \bar{f}(1 - (1 - \eta)|(\mathbb{E}_{\eta,\gamma}[A_{1,t}] - \mathbb{E}_{\eta,\gamma}[A_{2,t}])|) \\ f_{2,t+1} = \bar{f}(1 - \eta|(\mathbb{E}_{\eta,\gamma}[A_{2,t}] - \mathbb{E}_{\eta,\gamma}[A_{1,t}]])|) \end{cases}.$$

When $\eta = \frac{1}{2}$, $f_{1,t+1} = f_{2,t+1}$ so that $p_{1,t+1} = 1 - \frac{1}{2}f_{1,t+1}$ and $p_{2,t+1} = 1 - p_{1,t+1} = \frac{1}{2}f_{1,t+1}$.

The norm dynamics simplifies to

$$\begin{cases} x_{1,t+1} = (1 - \frac{1}{2}f_{1,t+1})((1-\lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]) + \frac{1}{2}f_{1,t+1}((1-\lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]) \\ x_{2,t+1} = \frac{1}{2}f_{1,t+1}((1-\lambda)x_{1,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{1,t}]) + (1 - \frac{1}{2}f_{1,t+1})((1-\lambda)x_{2,t} + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]) \end{cases} \tag{15}$$

and, adding up the two equations,

$$x_{1,t+1} + x_{2,t+1} = (1-\lambda)(x_{1,t} + x_{2,t}) + \lambda(\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]]). \tag{16}$$

Next we shall show that if $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (0, \frac{1}{2})$, or $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (0, \frac{1}{2})$, the term $\mathbb{E}_{\eta,\gamma}[A_{1,t}]] + \lambda\mathbb{E}_{\eta,\gamma}[A_{2,t}]$ depends only on the sum $z_t = x_{1,t} + x_{2,t}$, so that (16) can be used to characterize the dynamics of z_t.

Let us start from the region where $x_{1,t} \in (\frac{1}{2}, 1)$ and $x_{2,t} \in (0, \frac{1}{2})$. For $\mathbb{E}_{\eta,\gamma}[A_{1,t}]$ we have

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] = \gamma(\bar{b} < x_{1,t}) + \gamma(\bar{b} > x_{1,t} \ \wedge \ \bar{d} > x_{2,t})(1 - \frac{1}{4}) + \gamma(\bar{b} > x_{1,t} \wedge \bar{d} < x_{2,t})\frac{1}{2}.$$

Since payoffs are uniformly distributed in the region characterized by strategic substitutability

- $\gamma(\bar{b} < x_1) = 2x_1 - 1$
- $\gamma(\bar{b} > x_1 \wedge \bar{d} > x_2) = 4(1 - x_1)(\frac{1}{2} - x_2)$
- $\gamma(\bar{b} > x_1 \wedge \bar{d} < x_2) = 4(1 - x_1)x_2$

then

$$\begin{aligned} \mathbb{E}_{\eta,\gamma}[A_{1,t}] &= 2x_{1,t} - 1 + \frac{3}{2}(1 - x_{1,t})(1 - 2x_{2,t}) + 2(1 - x_{1,t})x_{2,t} \\ &= \frac{1}{2} - x_{2,t} + x_{1,t}(\frac{1}{2} + x_{2,t}) \end{aligned}$$

Turning to $\mathbb{E}_{\eta,\gamma}[A_{2,t}]$, we have

$$\mathbb{E}_{\eta,\gamma}[A_{2,t}] = \gamma(\bar{b} < x_{1,t} \ \wedge \ \bar{d} < x_{2,t})\frac{1}{4} + \gamma(\bar{b} > x_{1,t} \wedge \bar{d} < x_{2,t})\frac{1}{2}$$

so that, computing the probabilities of having norms within the given bounds,

$$\begin{aligned} \mathbb{E}_{\eta,\gamma}[A_{2,t}] &= 4\left(x_{1,t} - \frac{1}{2}\right)x_2\frac{1}{4} + 4(1 - x_{1,t})x_{2,t}\frac{1}{2} \\ &= \left(\frac{3}{2} - x_{1,t}\right)x_{2,t}. \end{aligned}$$

Importantly

$$\mathbb{E}_{\eta,\gamma}[A_{1,t}] + \mathbb{E}_{\eta,\gamma}[A_{2,t}] = \frac{1}{2}\left(1 + x_{1,t} + x_{2,t}\right).$$

Swapping the role of $x_{1,t}$ and $x_{2,t}$, we get the values of $\mathbb{E}_{\eta,\gamma}[A_{1,t}]$ and $\mathbb{E}_{\eta,\gamma}[A_{2,t}]$, and their sum, also in the region where $x_{1,t} \in (0, \frac{1}{2})$ and $x_{2,t} \in (\frac{1}{2}, 1)$.

Using the sum of average payoffs in (16), we obtain the dynamics for $z_t = x_{1,t} + x_{2,t}$:

$$z_{t+1} = (1-\lambda)z_t + \frac{\lambda}{2}(1+z_t) = \left(1 - \frac{\lambda}{2}\right) z_t + \frac{\lambda}{2} \tag{17}$$

The latter has a unique, and globally stable, steady state $z^* = 1$, implying that we can restrict the stability analysis of the norm dynamics on the line $x_1 + x_2 = 1$.

We turn to the analysis of the norm dynamics on the line $x_1 + x_2 = 1$. Without loss of generality we also impose $x_1 \in (\frac{1}{2}, 1)$. From the norms dynamics in (**??**) we obtain

$$x_{1,t+1} = (1 - \frac{1}{2} f_{1,t+1})[(1-\lambda)x_{1,t} + \lambda \mathbb{E}_{\eta,\gamma}[A_{1,t}]] + \frac{1}{2} f_{1,t+1}[(1-\lambda)(1-x_{1,t}) + \lambda \mathbb{E}_{\eta,\gamma}[A_{2,t}]]$$

where

$$\begin{cases} f_{1,t+1} = \bar{f}(1 - \frac{1}{2}|(\mathbb{E}_{\eta,\gamma}[A_{1,t}] - \mathbb{E}_{\eta,\gamma}[A_{2,t}])|) \\ \mathbb{E}_{\eta,\gamma}[A_{1,t}] = \frac{5}{2}x_{1,t} - \frac{1}{2} - x_{1,t}^2 \\ \mathbb{E}_{\eta,\gamma}[A_{2,t}] = \left(\frac{3}{2} - x_{1,t}\right)(1 - x_{1,t}). \end{cases}$$

Substituting and simplifying we obtain

$$x_{1,t+1} = \frac{1}{2} + \left(x_{1,t} - \frac{1}{2}\right)(1 + \lambda - \lambda x_{1,t})\left(1 - \bar{f}\left(1 - (2-x)\left(x - \frac{1}{2}\right)\right)\right)$$

or, defining $y_t = x_{t+1} - \frac{1}{2}$,

$$y_{t+1} = y_t \left(1 + \lambda\left(\frac{1}{2} - y_t\right)\right)\left(1 - \bar{f} + \bar{f} y_t \left(\frac{3}{2} - y_t\right)\right) = f(y_t) = y_t h(y_t). \tag{18}$$

The dynamics of y_t has $y^* = 0$ as steady state, whose stability can be assessed by looking at

$$f'(y)\big|_{y=0} = (h(y) + yh'(y))\big|_{y=0} = h(0) = \left(1 + \frac{\lambda}{2}\right)\left(1 - \bar{f}\right).$$

Imposing $f'(0) < 1$, we derive that the steady state $y^* = 0$, and thus $(x_1^*, x_2^*) = \left(\frac{1}{2}, \frac{1}{2}\right)$, is asymptotically stable when[25]

$$\bar{f} > \frac{\lambda}{2+\lambda}.$$

Note that instead

$$\bar{f} < \frac{\lambda}{2+\lambda} \tag{19}$$

implies that $y^* = 0$, and thus $(x_1^*, x_2^*) = \left(\frac{1}{2}, \frac{1}{2}\right)$, is unstable. The existence of other steady states and their stability depends on whether there exists $y^* \in \left(0, \frac{1}{2}\right)$ such that

$$f(y^*) = 0 \Leftrightarrow h(y^*) = 1 \quad \text{and} \quad |f'(y^*)| < 1 \Leftrightarrow |h(y^*) - y^* h'(y^*)| < 1 \quad .$$

Expliciting $h(y)$ from (18) we obtain the following third degree polynomial

$$h(y) = \bar{f}\lambda y^3 - \frac{3}{2}\bar{f}\left(1 + \frac{3}{2}\lambda\right)y^2 + \left(\frac{3}{2}\bar{f} + \frac{7}{4}\bar{f}\lambda - \lambda\right)y + \left(1 + \frac{\lambda}{2}\right)(1 - \bar{f}). \tag{20}$$

We shall use the properties of the function h and of its derivative h' in the interval $\left[0, \frac{1}{2}\right]$ to verify the presence of asymptotically stable fixed point. Having

$$\lim_{y \to \pm\infty} h(y) = \pm\infty \quad \text{and} \quad h\left(\frac{1}{2}\right) < 0$$

implies that there exists at most two solutions of $h(y) = 0$ in the interval $\left[0, \frac{1}{2}\right]$. We shall show that the exact number of solutions (steady states), and their stability, depends on the relative strength of λ and $\bar{f}$.

Let us consider first the case $\bar{f} < \frac{\lambda}{2+\lambda}$, so that $y^* = 0$ in unstable, as we derived in

[25] Alternatively one could derive the Jacobian of (15) in $\left(\frac{1}{2}, \frac{1}{2}\right)$. Irrespectively from the region where average actions are computed the result is

$$\boldsymbol{J}\left(\frac{1}{2}, \frac{1}{2}\right) = \begin{bmatrix} 1 - \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) & \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) - \frac{1}{2}\lambda \\ \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) - \frac{1}{2}\lambda & 1 - \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) \end{bmatrix}.$$

Calling μ_1 and μ_2 the eigenvalues of $J(x_t)$ we get

$$\mu_1 = 1 - \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) - 0.25\sqrt{\bar{f}^2(2+\lambda)^2 - \bar{f}(8+4\lambda)\lambda + 4\lambda^2}$$

$$\mu_2 = 1 - \bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) + 0.25\sqrt{\bar{f}^2(2+\lambda)^2 - \bar{f}(8+4\lambda)\lambda + 4\lambda^2}$$

While $\mu_1 < 1$ is always true, for $\mu_2 < 1$ we should verify that

$$-\bar{f}\left(\frac{1}{2} + \frac{1}{4}\lambda\right) + 0.25\sqrt{\bar{f}^2(2+\lambda)^2 - \bar{f}(8+4\lambda)\lambda + 4\lambda^2} < 0,$$

which is true when

$$\lambda < \frac{2\bar{f}}{1-\bar{f}} \;\Leftrightarrow\; \bar{f} > \frac{\lambda}{2+\lambda}.$$

(19). It can be verified that

$$h(0) > 1,$$

implying that there exists one and only solution $\overset{\diamond}{y}$ of $h(y) = 0$ in the interval $\left[0, \frac{1}{2}\right]$. The fact that such a fixed point is globally stable is implied by $y^* = 0$ being unstable and by the fact that $f'(y) > -1$ for all $y \in \left[0, \frac{1}{2}\right]$ (the latter is implied by $f'''(y) < 0$ and $f'\left(\frac{1}{2}\right) > -1$), so that $f'(\overset{\diamond}{y}) \in (-1, 1)$. We have shown that for $\bar{f} < \frac{\lambda}{2+\lambda}$ the norm dynamics converges to a $(\overset{\diamond}{x}_1, \overset{\diamond}{x}_2) \in \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$ for all initial conditions in $\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$.

Let us now turn to the case $\bar{f} = \frac{\lambda}{2+\lambda}$. Here, $h(0) = 1$. If $\lambda > \frac{2}{3}$ also $h'(0) > 0$ so that there exists two steady states, $y^* = 0$ and $\overset{\diamond}{y} \in \left(0, \frac{1}{2}\right)$, of which only the latter is stable (because $f'(0) = 0$ and $f'(y) > -1$ for all $y \in \left[0, \frac{1}{2}\right]$). Lower values of λ, holding fixed $\bar{f} = \frac{\lambda}{2+\lambda}$, imply instead $h'(0) \leq 0$ so that the only steady state is $y^* = 0$ and it is globally stable.

We turn now to $\bar{f} > \frac{\lambda}{2+\lambda}$, implying that $y^* = 0$ is asymptotically stable. Here, $h(0) < 1$. If $\bar{f}$ is close to $\frac{\lambda}{2+\lambda}$ and λ high enough, there exist two solutions of $h(y) = 1$ in the interval $\left(0, \frac{1}{2}\right)$. Of these solutions the largest, $\overset{\diamond}{y}$, is asymptotically stable. The norm dynamics has thus two locally stable fixed points $\left(\frac{1}{2}, \frac{1}{2}\right)$ and $(\overset{\diamond}{x}_1, \overset{\diamond}{x}_2) \in \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$. Increasing $\bar{f}$ above a threshold $\hat{f}$, threshold that depends on λ, implies instead that there are no solutions of $h(y) = 1$ (both $(h(0) - 1)$ and $h'(0)$ become rather negative) so that $y^* = 0$, corresponding to the norms $\left(\frac{1}{2}, \frac{1}{2}\right)$, is the only, globally stable, fixed point.

By symmetry we can repeat the reasoning when $x_{1,t} \in \left[0, \frac{1}{2}\right]$ and $x_{2,t} \in \left[\frac{1}{2}, 1\right]$ and find: converging to $(\mathring{x}_1, \mathring{x}_2)$ for low $\bar{f}$; convergence to $\left(\frac{1}{2}, \frac{1}{2}\right)$ for high $\bar{f}$; and convergence to either steady state, depending on the initial conditions, for intermediate $\bar{f}$. Since the dynamics is symmetric to the one found above the threshold $\bar{f}$ is the same.

□

A.8 Proofs of Section 3.2

Before we discuss the possible steady states in both complement and substitute environments for a singular payoff distribution, we provide a general convergence result. The next proposition ensures the for all parameters and every initial conditions, the norm dynamics converges and, moreover, all steady states that exist are asymptotically stable. Thus, we can partition the whole state space of norms in basins of attraction.

Proposition 6 *For all $\eta \in (0, 1)$, $\bar{f} \in (0, 1)$, $\lambda \in (0, 1)$, and $(\bar{b}, \bar{d}) \in [0, 1] \times [0, 1]$, if all material payoffs are determined by $(\bar{b}, \bar{d})$, then all steady states are asymptotically stable and their basins of attraction form a partition of the state space of norms and socialization levels.*

Proof

First note that the dynamics of $\boldsymbol{f}$ in (7) depends on norms through average payoffs. Moreover, when all players face the same game, average payoffs are constant for norms in the regions identified by Corollary 2.1. Thus, horizontal socializations $\boldsymbol{f}$ are constant in each region identified by Corollary 2.1, and equal to $\boldsymbol{f}^*$, and we can concentrate on the dynamics of norms. For $\boldsymbol{f}^*$ given, the dynamics of norms in (9) becomes

$$\mathbf{x}_{t+1} = \boldsymbol{\upsilon}(\zeta(\mathbf{x}_t, \boldsymbol{f}^*), \psi(\boldsymbol{f}^*)).$$

or, expliciting the functions $\boldsymbol{\upsilon}$ and ζ,

$$\mathbf{x}_t = (1-\lambda)\begin{bmatrix} p_{1,t}(\boldsymbol{f}^*) & 1 - p_{1,t}(\boldsymbol{f}^*) \\ p_{2,t}(\boldsymbol{f}^*) & 1 - p_{2,t}(\boldsymbol{f}^*) \end{bmatrix} \mathbf{x}_{t-1} + constant,$$

where the constant depends on average payoffs and $\boldsymbol{f}^*$ and the expression of $p_{1,t}$ and $p_{2,t}$ is in Corollary 1.1.

The norm dynamics within each region of the state space depends linearly on previous norms through a stochastic matrix, whose composition is also stochastic. We can thus solve the norm dynamics and get

$$\mathbf{x}_t - \mathbf{x}^* = (1-\lambda)^t \mathbf{P}(t)(\mathbf{x}_0 - \mathbf{x}^*),$$

where $\mathbf{P}(t)$ is a stochastic matrix and $\mathbf{x}^*$ is a steady states. Taking the limit, and assuming that the steady state $\mathbf{x}^*$ is in the region considered, we have convergence for all $\mathbf{x}_0$ within the same region (and possibly outside from it if there are no steady state in other regions) Below, we characterize all the possible steady states of (9) and their possible basin of attraction.

□

A.9 Point Distribution: Steady States

In this section, we discuss all the possible steady state with point distribution.

We define e as a generic element of $\mathcal{E}$, thus we can enumerate the possible steady states as $\boldsymbol{e}^1, \boldsymbol{e}^2, ...$

Proposition 7 *[Assimilation and Norm Neutrality]*
Consider the norm and socialization level dynamics in (9). For all $\eta \in (0,1)$, $\bar{b} \in (0,1)$, $\bar{d} \in (0,1)$, $\bar{f} \in (0,1)$, $\lambda \in (0,1)$

- $\boldsymbol{e}^1 = (1,1)$ *and* $\boldsymbol{e}^2 = (0,0) \in \mathcal{E}$. *In both steady states the socialization level is* $f^* = (\bar{f}, \bar{f})$ *and the average actions are, respectively,* $\mathbb{E}_\eta[\boldsymbol{A}^*] = (1,1)$ *and* $\mathbb{E}_\eta[\boldsymbol{A}^*] = (0,0)$,

- $\boldsymbol{e}^3 = \left(\frac{1}{2}, \frac{1}{2}\right) \in \mathcal{E}$ *if and only if the original* 2×2 *symmetric game has multiple Nash equilibria. At the steady state the socialization level is* $f^* = (\bar{f}, \bar{f})$ *and the average action is* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(\frac{1}{2}, \frac{1}{2}\right)$

Proposition 8 ***[Polarization]***
Consider the norm and socialization level dynamics in (9). For all $\eta \in (0,1)$, $\bar{b} \in (0,1)$, $\bar{d} \in (0,1)$, $\bar{f} \in (0,1)$, $\lambda \in (0,1)$

- $\boldsymbol{e}^4 = (\phi_1, \phi_2, \bar{f}\eta, \bar{f}(1-\eta)) \in \mathcal{E}$ *if and only if* $\phi_1 > max\{\bar{d}, \bar{b}\} \wedge \phi_2 < min\{\bar{d}, \bar{b}\}$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = (1, 0)$.
- $\boldsymbol{e}^5 = (1-\phi_1, 1-\phi_2, \bar{f}\eta, \bar{f}(1-\eta)) \in \mathcal{E}$ *if and only if* $1-\phi_1 < max\{\bar{d}, \bar{b}\} \wedge 1-\phi_2 > min\{\bar{d}, \bar{b}\}$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = (0, 1)$.

Proposition 9 *[Integration Without Assimilation]*
Consider the norm and socialization level dynamics in (9). For all $\eta \in (0,1)$, $\bar{b} \in (0,1)$, $\bar{d} \in (0,1)$, $\bar{f} \in (0,1)$, $\lambda \in (0,1)$, *with* $\bar{b} < \bar{d}$

- $\boldsymbol{e}^6_{\bar{d}} = \left(1 - \frac{1}{2}(1-\eta)(1-\phi_1), 1 - \frac{1}{2}(1-\phi_2)(1-\eta), \bar{f}\left(1 - \frac{1}{2}(1-\eta)^2\right), \bar{f}\left(1 - \frac{1}{2}\eta(1-\eta)\right)\right) \in \mathcal{E}$ *if and only if* $1 - \frac{1}{2}(1-\phi_2)(1-\eta) < \bar{d} < 1 - \frac{1}{2}(1-\eta(1-\phi_1))$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(1, \frac{1}{2}(1+\eta)\right)$.
- $\boldsymbol{e}^7_{\bar{d}} = \left(1 - \frac{1}{2}\eta\phi_1, 1 - \frac{1}{2}\eta\phi_2, \bar{f}\left(1 - \frac{1}{2}\eta(1-\eta)\right), \bar{f}\left(1 - \frac{1}{2}\eta^2\right)\right) \in \mathcal{E}$ *if and only if* $1 - \frac{1}{2}\eta\phi_1 < \bar{d} < 1 - \frac{1}{2}\eta\phi_2$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(1 - \frac{1}{2}\eta, 1\right)$.
- $\boldsymbol{e}^8_{\bar{d}} = \left(\frac{1}{2}\eta\phi_1, \frac{1}{2}\eta\phi_2, \bar{f}\left(1 - \frac{1}{2}\eta(1-\eta)\right), \bar{f}\left(1 - \frac{1}{2}\eta^2\right)\right) \in \mathcal{E}$ *if and only if* $\frac{1}{2}\eta\phi_2 < \bar{b} < \frac{1}{2}\eta\phi_1$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(\frac{1}{2}\eta, 0\right)$.
- $\boldsymbol{e}^9_{\bar{d}} = \left(\frac{1}{2}(1-\eta)(1-\phi_1), \frac{1}{2}(1-\eta)(1-\phi_2), \bar{f}\left(1 - \frac{1}{2}(1-\eta)^2\right), \bar{f}\left(1 - \frac{1}{2}\eta(1-\eta)\right)\right) \in \mathcal{E}$ *if and only if* $\frac{1}{2}(1-\eta)(1-\phi_1) < \bar{b} < \frac{1}{2}(1-\eta)(1-\phi_2)$. *The actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(0, \frac{1}{2}(1-\eta)\right)$.

Proposition 10 *[Partial Polarization]*
Consider the norm and socialization level dynamics in (9). For all $\eta \in (0,1)$, $\bar{b} \in (0,1)$, $\bar{d} \in (0,1)$, $\bar{f} \in (0,1)$, $\lambda \in (0,1)$, *with* $\bar{d} < \bar{b}$

- $\boldsymbol{e}^6_{\bar{b}} = \left(1 - \frac{1}{2}(1-\phi_1)(1+\eta), 1 - \frac{1}{2}(1-\phi_2)(1+\eta), \bar{f}(\frac{1}{2}(1+\eta^2)), \bar{f}(1 - \frac{1}{2}\eta(1+\eta))\right) \in \mathcal{E}$ *if and only if* $\bar{b} < 1 - \frac{1}{2}(1-\phi_1)(1+\eta) \wedge \bar{d} < 1 - \frac{1}{2}(1-\phi_2)(1+\eta)$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(1, \frac{1}{2}(1-\eta)\right)$.
- $\boldsymbol{e}^7_{\bar{b}} = \left(1 - \phi_1(1 - \frac{1}{2}\eta), 1 - \phi_2(1 - \frac{1}{2}\eta), \bar{f}(1 - \frac{1}{2}\eta(1-\eta)), \bar{f}(\frac{1}{2}\eta(3-\eta))\right) \in \mathcal{E}$ *if and only if* $\bar{d} < 1 - \phi_1(1 - \frac{1}{2}\eta) \wedge \bar{b} < 1 - \phi_2(1 - \frac{1}{2}\eta)$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(\frac{1}{2}\eta, 1\right)$.
- $\boldsymbol{e}^8_{\bar{b}} = \left(\phi_1(1 - \frac{1}{2}\eta), \phi_2(1 - \frac{1}{2}\eta), \bar{f}(1 - \frac{1}{2}\eta(1-\eta)), \bar{f}(\frac{1}{2}\eta(3-\eta))\right) \in \mathcal{E}$ *if and only if* $\phi_2(1 - \frac{1}{2}\eta) < \bar{d} < \phi_1(1 - \frac{1}{2}\eta) \wedge \bar{b} > \phi_1(1 - \frac{1}{2}\eta)$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(1 - \frac{1}{2}\eta, 0\right)$.

- $\boldsymbol{e}_{\bar{b}}^{9} = \left(\frac{1}{2}(1+\eta)(1-\phi_1), \frac{1}{2}(1+\eta)(1-\phi_2), \bar{f}(\frac{1}{2}(1+\eta^2)), \bar{f}(1-\frac{1}{2}\eta(1+\eta))\right) \in \mathcal{E}$ *if and only if* $\frac{1}{2}(1+\eta)(1-\phi_1) < \bar{d} < \frac{1}{2}(1+\eta)(1-\phi_2) \wedge \bar{b} > \frac{1}{2}(1+\eta)(1-\phi_2)$. *The average actions at the steady state are* $\mathbb{E}_\eta[\boldsymbol{A}^*] = \left(0, \frac{1}{2}(1+\eta)\right)$.

Corollary 10.1 *The steady state described in Proposition 7, 8, 9, and 10 are the only possible steady states of* (9). *Moreover*

- *If the game has strategic complements,* $\bar{b} < \frac{1}{2} < \bar{d}$, (9) *has a minimum of three steady states,* $(\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3)$, *and a maximum of nine,* $(\boldsymbol{e}^1, ..., \boldsymbol{e}_{\bar{d}}^9)$.
- *If the game has strategic substitutes,* $\bar{b} < \frac{1}{2} < \bar{d}$, (9) *has a minimum of three steady states,* $(\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3)$, *and a maximum of nine,* $(\boldsymbol{e}^1, ..., \boldsymbol{e}_{\bar{b}}^9)$.

Proofs of Propositions 7-10 and Corollary 10.1

If the distribution γ is singular, then $\mathbb{E}_{\eta,\gamma}[\boldsymbol{A}] = \mathbb{E}_\eta[\boldsymbol{A}]$. In order to prove the previous propositions we need to substitute $\mathbb{E}_\eta[\boldsymbol{A}]$ in equation (9). Having constant payoffs when norms are within specific bound given by $\bar{b}$ and $\bar{d}$ and a linear dynamics of norms due to cognitive dissonance, (5), proves convergence and leads to the following basins of attraction $B(\boldsymbol{e}^1)$, $B(\boldsymbol{e}^2)$, ... defined as a function of norms regions in Corollary 2.1.

1. $$B(\boldsymbol{e}^1)\begin{cases} \ni R_{1_*,1_*} & \textit{always} \\ \ni R_{1_*,1} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^6 \in R_{1_*,1_*} \\ \ni R_{1,1_*} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^6 \in R_{1_*,1_*} \\ \ni R_{1_*,0_*} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}^4 \in R_{1_*,1} \wedge \boldsymbol{e}_{\bar{d}}^6 \in R_{1_*,1_*} \\ \ni R_{0_*,1_*} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}^5 \in R_{1,1_*} \wedge \boldsymbol{e}_{\bar{d}}^6 \in R_{1_*,1_*} \end{cases}$$

2. $$B(\boldsymbol{e}^2)\begin{cases} \ni R & \text{iff} \quad \boldsymbol{e}^2 \in R \\ \ni R_{1_*,0_*} & \text{iff} \quad \boldsymbol{e}^4 \in R \\ \ni R_{0_*,1_*} & \text{iff} \quad \boldsymbol{e}^5 \in R \\ \ni R_{1_*,1} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^6 \in R \\ \ni R_{1,1_*} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^7 \in R \\ \ni R_{0,0_*} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^8 \in R \\ \ni R_{0_*,0} & \text{iff} \quad \bar{d} > \bar{b} \wedge \boldsymbol{e}_{\bar{d}}^9 \in R \\ \ni R_{1_*,0} & \text{iff} \quad \bar{b} > \bar{d} \wedge \boldsymbol{e}_{\bar{b}}^6 \in R \\ \ni R_{0,1_*} & \text{iff} \quad \bar{b} > \bar{d} \wedge \boldsymbol{e}_{\bar{b}}^7 \in R \\ \ni R_{1,0_*} & \text{iff} \quad \bar{b} > \bar{d} \wedge \boldsymbol{e}_{\bar{b}}^8 \in R \\ \ni R_{0_*,1} & \text{iff} \quad \bar{b} > \bar{d} \wedge \boldsymbol{e}_{\bar{b}}^9 \in R \end{cases}$$

$$3.\ B(\boldsymbol{e}^3)\begin{cases}\ni R_{0_*,0_*} & \textit{always}\\ \ni R_{0,0_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^8_{\bar{d}}\in R_{0_*,0_*}\\ \ni R_{0_*,0} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^9_{\bar{d}}\in R_{0_*,0_*}\\ \ni R_{1_*,0_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^4\in R_{0,0_*}\wedge \boldsymbol{e}^8_{\bar{d}}\in R_{0_*,0_*}\\ \ni R_{0_*,1_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^5\in R_{0_*,0}\wedge \boldsymbol{e}^9_{\bar{d}}\in R_{0_*,0_*}\end{cases}$$

$$4.\ B(\boldsymbol{e}^4)\begin{cases}\ni R_{1_*,0_*} & \text{iff}\quad \boldsymbol{e}^4\in\mathcal{E}\\ \ni R_{1_*,0} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^6_{\bar{b}}\in R_{1_*,0_*}\\ \ni R_{1,0_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^8_{\bar{b}}\in R_{1_*,0_*}\end{cases}$$

$$5.\ B(\boldsymbol{e}^5)\begin{cases}\ni R_{0_*,1_*} & \text{iff}\quad \boldsymbol{e}^5\in\mathcal{E}\\ \ni R_{0,1_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^7_{\bar{b}}\in R_{0_*,1_*}\\ \ni R_{0_*,1} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^9_{\bar{b}}\in R_{0_*,1_*}\end{cases}$$

$$6.\ B(\boldsymbol{e}^6_{\bar{d}})\begin{cases}\ni R_{1_*,1} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^6_{\bar{d}}\in\mathcal{E}\\ \ni R_{1_*,0_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^4\in R_{1_*,1}\end{cases}$$

$$7.\ B(\boldsymbol{e}^7_{\bar{d}})\begin{cases}\ni R_{1,1_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^7_{\bar{d}}\in\mathcal{E}\\ \ni R_{0_*,1_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^5\in R_{1,1_*}\end{cases}$$

$$8.\ B(\boldsymbol{e}^8_{\bar{d}})\begin{cases}\ni R_{0,0_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^8_{\bar{d}}\in\mathcal{E}\\ \ni R_{1_*,0_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^4\in R_{0,0_*}\end{cases}$$

$$9.\ B(\boldsymbol{e}^9_{\bar{d}})\begin{cases}\ni R_{0_*,0} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^9_{\bar{d}}\in\mathcal{E}\\ \ni R_{0_*,1_*} & \text{iff}\quad \bar{d}>\bar{b}\wedge \boldsymbol{e}^5\in R_{0_*,0}\end{cases}$$

$$10.\ B(\boldsymbol{e}^6_{\bar{b}})\begin{cases}\ni R_{1_*,0} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^6_{\bar{b}}\in\mathcal{E}\\ \ni R_{1_*,0_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^4\in R_{1_*,0}\end{cases}$$

$$11.\ B(\boldsymbol{e}^7_{\bar{b}})\begin{cases}\ni R_{0,1_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^7_{\bar{b}}\in\mathcal{E}\\ \ni R_{0_*,1_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^5\in R_{0,1_*}\end{cases}$$

$$12.\ B(\boldsymbol{e}^8_{\bar{b}})\begin{cases}\ni R_{1,0_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^8_{\bar{b}}\in\mathcal{E}\\ \ni R_{1_*,0_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^4\in R_{1,0_*}\end{cases}$$

$$13.\ B(\boldsymbol{e}^9_{\bar{b}})\begin{cases}\ni R_{0_*,1} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^9_{\bar{b}}\in\mathcal{E}\\ \ni R_{0_*,1_*} & \text{iff}\quad \bar{b}>\bar{d}\wedge \boldsymbol{e}^5\in R_{0_*,1}\end{cases}$$

Considering all the possible steady states and basins of attraction and we can easily check that

$$\bigcup_{\forall \boldsymbol{e}^i \in E} B(\boldsymbol{e}^i) = [0,1]^2 \times [0,1]^2$$

Thus, independently on parameters $\bar{b}, \bar{d}, \eta, \lambda, \bar{f}$ the dynamics (9) converges to a stable steady state for all initial norms $\mathbf{x}_0$.

□

B Microfoundation of Parental Transmission

B.1 Flexibility Parameter Optimal Choice

In this section we shall show that the vector of flexibility parameters $f = (f_{1,t}, f_{2,t})$ can arise as the equilibrium of the game played by parents of groups $i = 1, 2$ who face the following optimization problem (see e.g. Panebianco, 2014)

$$\max_{1-f_{i,t} \in [0,1]} \left\{ -(\theta_{i,t} - x_{i,t})^2 - \frac{1}{2}(1 - f_{i,t})^2 \right\}. \tag{21}$$

Before we proceed, it is necessary to adapt the concept of *cultural substitution* to our framework. According to Bisin and Verdier (2001) there is *cultural substitution* whenever "parents have fewer incentives to socialize their children the more widely dominant are their values in the population". In our model the cultural traits are continuous and the communities are fixed, moreover, according to (5) the transmitted norm for the community i, $\theta_{i,t}$, depends on the average action played. For his reason we are interested in re-define the concept of *cultural substitution* (*complementarity*) with respect the difference between the actions of community i and those of the whole society.

To be consistent with the standard literature about cultural transmission, we have written the maximization problem (21) with respect the direct (vertical) socialization effort $1 - f_{i,t}$. Since we are interested in the horizontal socialization, we study *cultural substitution* (*complementarity*) with respect $f_{i,t}$.

Recall that the average action in the whole society is

$$\bar{A}_t = \eta \mathbb{E}_{\eta,\gamma}[A_{1,t}] + (1 - \eta)\mathbb{E}_{\eta,\gamma}[A_{2,t}].$$

Definition 1 *$f_{i,t}$ satisfies the cultural substitution (complementarity) property if, for all parameter values, is a continuous, strictly decreasing (increasing) function with respect* $|\mathbb{E}_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|$

We define cultural substitution (complementarity) with respect the average actions and not with respect population shares η and $1-\eta$. This distinction is important and it is due to the fact that in our framework population shares are fixed, while norms and actions evolve over time. Moreover agents belonging to different *ex-ante* community may end up to have the same norm and play the same action.

We further assume that parents are able to anticipate the offsprings' future choices in the young age, (3), buy that are not able to anticipate their payoffs in the adult age.

Proposition 11
Given the parents problem (21), there exists an $f_{i,t}$ that satisfies the cultural substitution property.

Proof of Proposition 11

First we change variable and set $\tau_{i,t} = 1 - f_{i,t}$ so that (21) becomes

$$\max_{\tau_{i,t}\in[0,1]} \left\{ -(\theta_{i,t} - x_{i,t})^2 - \frac{1}{2}\tau_{i,t}^2 \right\}.$$

Parents are aware that

$$x_{i,t} = p_i(\tau_t)\theta_{i,t} + (1 - p_i(\tau_t))\theta_{-i,t}.$$

Thus they solve

$$\max_{\tau_{i,t}\in[0,1]} \left\{ -(\theta_{i,t} - p_i(\tau_t)\theta_{i,t} - (1 - p_i(\tau_t))\theta_{-i,t})^2 - \frac{1}{2}\tau_{i,t}^2 \right\},$$

$$\max_{\tau_{i,t}\in[0,1]} \left\{ -(1 - p_i(\tau_t))^2(\theta_{i,t} - \theta_{-i,t})^2 - \frac{1}{2}\tau_{i,t}^2 \right\}.$$

Solving the first order condition

$$\frac{\partial U}{\partial \tau_{i,t}} = 0 : \qquad 2\frac{\partial p_i(\theta_t)}{\partial \tau_{i,t}}(1 - p_i(\theta_t))(\theta_{i,t} - \theta_{-i,t})^2 - \tau_{i,t} = 0$$

where

$$p_i(\tau_t) = \frac{\tau_{i,t}(1 - (1 - \tau_{-i,t})(1 - \eta))}{1 - \tau_{i,t}\eta - (1 - \tau_{-i,t})(1 - \eta)} = \frac{\tau_{i,t}(\eta + \tau_{-i,t}(1 - \eta))}{\tau_{i,t}\eta + \tau_{-i,t}(1 - \eta)},$$

$$\frac{\partial p_i(\theta_t)}{\partial \tau_{i,t}} = \frac{\tau_{-i,t}(1 - \eta)(\eta + \tau_{-i,t}(1 - \eta)))}{(\tau_{i,t}\eta + \tau_{-i,t}(1 - \eta))^2},$$

$$1 - p_i(\tau_t) = \frac{(1 - \tau_{i,t})(1 - \eta)\tau_{-i,t}}{\tau_{i,t}\eta + \tau_{-i,t}(1 - \eta)}.$$

The latter imply

$$2\frac{\tau_{-i,t}^2(1 - \tau_{-i,t})(1 - \eta)^2(\eta + \tau_{-i,t}(1 - \eta)))}{(\tau_{i,t}\eta + \tau_{-i,t}(1 - \eta))^3}(\theta_{i,t} - \theta_{-i,t})^2 - \tau_{i,t} = 0. \tag{22}$$

(from (22) it is also evident that $\frac{\partial^2 U}{\partial \tau_i^2} < 0$, so that the second order condition is satisfied).

Next, we check the number of solutions of (22) or, equivalently, the solutions of

$$\tau_{i,t} = 2(\theta_{i,t} - \theta_{-i,t})^2 \tau_{-i,t}^2 (1-\eta)^2 (\eta + \tau_{-i,t}(1-\eta)) \frac{(1-\tau_{i,t})}{(\tau_{i,t}\eta + \tau_{-i,t}(1-\eta))^3} \tag{23}$$

On the left hand side of (23) we have a linear function while on the right hand side we have a decreasing function (this is evident by computing the derivative of the r.h.s. with respect to $\tau_{i,t}$). Thus for each $\tau_{-i,t}$ there exists only one real and positive best reply $\tau_{i,t}$. Since we have the two best reply functions (one for each group of parents) we can use the Brouwer's fixed point theorem to ensure that the game where parents solve (21) admits at least an equilibrium in the space $[0, 1]$.

If $\tau_{-i,t} = 0$, then $\tau_{i,t} = 0$ implying that $(0, 0)$ is always a solution of (23). Next we shall prove that also an interior equilibrium exists. Sufficient conditions are that $\tau_{i,t} < 1$ when $\tau_{-i,t} = 1$ for both $i = 1, 2$ and that best replies in zero have first derivative higher than 1.

We first show that if $\tau_{-i,t} = 1$, then $\tau_{i,t} < 1$. Substituting $\tau_{-i,t} = 1$ in (23) we obtain

$$\tau_{i,t} = 2 \cdot (\theta_{i,t} - \theta_{-i,t})^2 \cdot (1-\eta)^2(\eta + (1-\eta)) \cdot \frac{(1-\tau_{i,t})}{(\tau_{i,t}\eta + (1-\eta))^3}.$$

whose solution in $\tau_{i,t}$ can only be in the interval $(0, 1)$.

Next we use implicit function theorem on (22), which defines the function F, to show that best replies in zero are increasing with first derivative higher than 1. We obtain

$$\tau'_{i,t}(\tau_{-i,t}) = -\frac{F'_{\tau_{-i,t}}}{F'_{\tau_{i,t}}},$$

$$F'_{\tau_{-i,t}} = \frac{(1-\eta)^2\eta\tau_{-i,t}(1-\tau_{i,t})((1-\eta)\tau_{-i,t}(3\tau_{i,t}-1) + 2\eta\tau_{i,t})}{(\tau_{i,t}\eta + \tau_{-i,t}(1-\eta))^4}(\theta_{i,t} - \theta_{-i,t})^2,$$

$$F'_{\tau_{i,t}} = -\frac{(1-\eta)^2\tau_{-i,t}^2(\eta(\tau_{-i,t}-1) - \tau_{-i,t})(\eta(\tau_{-i,t} - 2\tau_{i,t} - 3) - \tau_{-i,t})}{(\tau_{i,t}\eta + \tau_{-i,t}(1-\eta))^4}(\theta_{i,t} - \theta_{-i,t})^2 - 1.$$

Defining $\Delta\theta_t = \theta_{i,t} - \theta_{-i,t}$

$$\tau'_{i,t}(\tau_{-i,t}) = \frac{(1-\eta)^2\eta\tau_{-i,t}(1-\tau_{i,t})((1-\eta)\tau_{-i,t}(3\tau_{i,t}-1) + 2\eta\tau_{i,t})\Delta\theta_t^2}{(1-\eta)^2\tau_{-i,t}^2(\eta(\tau_{-i,t}-1) - \tau_{-i,t})(\eta(\tau_{-i,t} - 2\tau_{i,t} - 3) - \tau_{-i,t})\Delta\theta_t^2 - (\tau_{i,t}\eta + \tau_{-i,t}(1-\eta))^4}. \tag{24}$$

Although we cannot use implicit function theorem in the point $\tau_{-i,t} = 0$, we can study $\tau'_{i,t}(\tau_{-i,t})$ when $\tau_{-i,t} \to 0$.

From (23), $\tau_{i,t}$ is the solution of a quartic equation, thus $\tau_{i,t} = o(\tau^{\alpha}_{-i,t})$. Therefore

$$\tau'_{i,t}(\tau_{-i,t}) = \frac{(1-\eta)^2\eta\tau_{-i,t}(1-o(\tau^{\alpha}_{-i,t}))((1-\eta)\tau_{-i,t}(3o(\tau^{\alpha}_{-i,t})-1) + 2\eta o(\tau^{\alpha}_{-i,t}))\Delta\theta_t^2}{(1-\eta)^2\tau_{-i,t}^2(\eta(\tau_{-i,t}-1) - \tau_{-i,t})(\eta(\tau_{-i,t} - 2o(\tau^{\alpha}_{-i,t}) - 3) - \tau_{-i,t})\Delta\theta_t^2 - (o(\tau^{\alpha}_{-i,t})\eta + \tau_{-i,t}(1-\eta))^4} \tag{25}$$

$$\lim_{\tau_{-i,t}\to 0} \tau'_{i,t}(\tau_{-i,t}) = \lim_{\tau_{-i,t}\to 0} \frac{\tau_{-i,t}(\tau_{-i,t} + \tau^{\alpha}_{-i,t})}{\tau^3_{-i,t}(-\tau_{-i,t} - \tau^{\alpha}_{-i,t}) - \tau^{4\alpha}_{-i,t} - \tau^4_{-i,t}}$$

In studying $\lim_{\tau_{-i,t}} \tau'_{i,t}(\tau_{-i,t})$ we have to consider, both at the numerator and the denominator, only the smallest power of $\tau_{-i,t}$, thus we have to distinguish different cases.

- If $\alpha > 1$ then

  $$\lim_{\tau_{-i,t}\to 0} \frac{-\tau^2_{-i,t}}{-\tau^4_{-i,t}} = +\infty,$$

  which is not possible, since it is not consistent with $\tau_{i,t} = o(\tau^{\alpha}_{-i,t})$.

- If $0 < \alpha < 1$ then

  $$\lim_{\tau_{-i,t}\to 0} \frac{-\tau^{1+\alpha}_{-i,t}}{-\tau^{4\alpha}_{-i,t}} = +\infty \qquad \text{if} \ \ \alpha > \frac{1}{3},$$

  so that $\alpha > \frac{1}{3}$ is not consistent with $\tau_{i,t} = o(\tau^{\alpha}_{-i,t})$.

We can conclude that it must be $\frac{1}{3} < \alpha < 1$. Both the numerator and the denominator of (25) go to zero when $\tau_{-i,t} \to 0$ and, since the denominator is a polynomial of higher order in $\tau_{-i,t}$, the slope of the best-reply function is unbounded (from above) in a sufficiently small neighborhood of zero.

We conclude the proof by showing that that best replies positively depend on $|\mathbb{E}_{\eta,\gamma}[A_{i,t-1}] - \mathbb{E}_{\eta,\gamma}[A_{-i,t-1}]|$. This is sufficient to show that the socialization problem satisfy the cultural substitutions properties.

Defining $\Delta\mathbb{E}_{\eta,\gamma}[A_{t-1}] = \mathbb{E}_{\eta,\gamma}[A_{i,t-1}] - \mathbb{E}_{\eta,\gamma}[A_{-i,t-1}]$ and substituting equation (5) we get

$$\tau_{i,t} = 2((1-\lambda)(x_{i,t-1} - x_{-i,t-1}) + \lambda\Delta\mathbb{E}_{\eta,\gamma}[A_{t-1}])^2 \tau^2_{-i,t}(1-\eta)^2(\eta + \tau_{-i,t}(1-\eta))\frac{(1-\tau_{i,t})}{(\tau_{i,t}\eta + \tau_{-i,t}(1-\eta))^3}$$

$\tau_{i,t}$ positively depend on $\mathbb{E}_{\eta,\gamma}[A_{i,t-1}] - \mathbb{E}_{\eta,\gamma}[A_{-i,t-1}]$. Since $|\mathbb{E}_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|$ is equivalent to $|(\mathbb{E}_{\eta,\gamma}[A_{i,t-1}] - \mathbb{E}_{\eta,\gamma}[A_{-i,t-1}])(1-\eta)|$, then $\tau_{i,t}$ positively depends on $|E_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|$. We also know that $f_{i,t} = 1 - \tau_{i,t}$. Therefore, given the parents problem (21), $f_{i,t}$ satisfies the cultural substitution property. The qualitative result on socialization level of our rule (7) is equivalent to the one derived in a rational transmission model that stems from preference for offspring with similar cultural traits.

In Figure 7, we show the change of best replies, and thus of the interior equilibrium, after a change of $|\mathbb{E}_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|$ for both $i = 1, 2$. In particular, if $|\mathbb{E}_{\eta,\gamma}[A_{i,t}] - \bar{A}_t|$ decreases also the equilibrium vertical socialization decreases. Although we represent the case with only one interior equilibrium, the same applies with more equilibria if one selects the

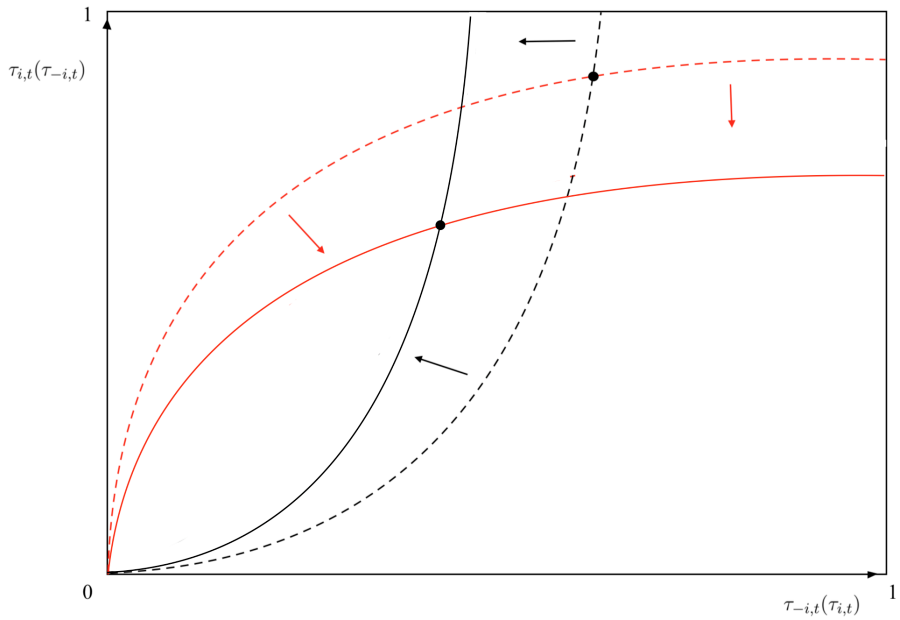


Figure 7: Change of best replies of the Parental Transmission Game, and of the related Nash Equilibrium, after a decrease in average payoffs difference.

equilibrium with max or min value of vertical socialization.

□