On the Joint Evolution of Culture and Institutions: Elites and Civil Society*

Alberto Bisin†

New York University, NBER, and CEPR

alberto.bisin@nyu.edu

Thierry Verdier‡

Paris School of Economics and Ecole des Ponts-Paris Tech, PUC-Rio and CEPR

tv@pse.ens.fr

September 29, 2021

Abstract

In this paper we provide an abstract modeling of the interaction between culture and institutions and their effects on economic variables of interest, notably, e.g., long-run economic activity. We characterize conditions on the socio-economic environment such that culture and institutions complement (resp. substitute) each other, giving rise to a *multiplier* effect which amplifies (resp. dampens) their combined ability to spur socio-economic activity. We show how the joint dynamics of culture and institutions may display interesting non-ergodic behavior, hysteresis, oscillations, depending on the form of the interaction between culture and institutions. The model can be specialized to study the political economy of elites and civil society for the determination of long-run socio-economic activity in different contexts. We illustrate this by studying the transition away from *extractive institutions* and the formation of *civic capital* in two example societies.

First draft: September 2015

Journal of Economic Literature Classification Numbers: O10, P16, P48.

Keywords: Cultural transmission; evolution of institutions.

*Thanks to Alberto Alesina, Gary Becker, Jess Benhabib, Georgy Egorov, Martin Fiszbein, Luigi Guiso, David Levine, Dilip Mookherjee, Massimo Morelli, Nicola Persico, Louis Putterman, Kenneth Shepsle, Shlomo Weber, and the many others for comments. Thanks to Simone Meraglia for historical references. We have presented the paper widely before and after writing it and we thank all the audiences for help, suggestions, and support. Earlier drafts of this paper circulated under the title, "On the Joint Evolution of Culture and Institutions." This research has been undertaken under financial support from the ERC Advanced Grant TECTACOM °324004.

†Department of Economics, New York University, 19 West Fourth Street, New York, NY, 10012, USA.

‡Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France.

1 Introduction

"Era questo un ordine buono, quando i cittadini erano buoni [...] ma diventati i cittadini cattivi, divento' tale ordine pessimo." [This was a good institutional order when citizens were good [...] but when citizen became bad, it turned into an horrible order; our translation]; Niccolo' Machiavelli, *Discorsi,* I. 16, 1531

[...] "among a people generally corrupt, liberty cannot long exist." Edmund Burke, Letter to the Sheriffs of Bristol (1777-04-03).

"If there be no virtue among us, no form of government can render us secure. To suppose that any form of government will secure liberty or happiness without any virtue in the people is an illusion." James Madison, 20 June 1788, Papers 11:163

The distribution of income across countries in the world is very unequal. A thriving literature in economics and political science studies which factors may account for long-run income inequality. In this context, institutions and culture are often run against each other as possible explanatory factors. In this paper we focus instead on the interactions between culture and institutions, under the understanding that these interactions drive important historical processes which in turn affect economic activity and, more generally, socio-economic prosperity over the long run.

For instance, the formation of inclusive institutions protecting property rights in England has arguably spurred economic activity after the development of an appropriate system of ideas and beliefs, the "bourgeois Ideology,"in McCloskey (2006, 2010, 2017) and the "Industrial Enlightment" in Mokyr (2016), which fundamentally contributed to laying the grounds for the advent of the Industrial Revolution (see also Doepke and Zilibotti, 2008). Relatedly, the institutional independence from the Sacred Roman Empire obtained by several Italian cities ("Communes") in the Middle Ages has had significant and very persistent effects on their economic prosperity, arguably also through the development of the stock of civic capital of their citizens, transmitted across generations and acquired by those who moved into these cities over the centuries (Guiso, Sapienza, and Zingales 2008, 2016).[1]

Abstracting from specific contextual instances, in this paper we provide a systematic modeling framework to analyze the role of the interaction of culture and institutions as determinants of economic variables of interest, notably, e.g., long run economic activity. The stylized dynamics

[1]Along similar lines, the Roman Empire's reliance on slavery institutions, which developed jointly and alongside its aristocracy's ideological stigma against manual occupations, crucially contributed to the fall of the empire in the 4th century CE (Schiavone, 1996). Economic activity after the end of colonial institutions in Africa also appears to have been modulated by the development of cultural traits and norms of behavior orginated during or before colonial times (Lowes and Monteiro, 2017, on the legacy of the rubber concessions in the Congo Free State; Lowes, Nunn, Robinson, and Weigal, 2017, on the Kuba Kingdom). Related references include Gorodnichenko and Roland (2017) on market institutions and individualism, Greif and Tabellini (2010, 2017) on norms of kinship and moral systems, Boranbay and Guerriero (2019) on inclusive institutions and culture of cooperation.

we obtain as a result map into novel tools for the empirical analysis of the effects of exogenous variations in cultural and/or institutional phenomena. This modeling framework can be directly specialized to study the political economy of elites and civil society in different socio-economic contexts, illustrating various interesting regularities through economic history, like e.g., the formation and circulation of the elites, the transition away from extractive institutions, the accumulation of civic capital, the reliance of autocratic regimes on religious legitimacy, and their effects on economic activity.

The modeling of the joint evolution of culture and institutions we provide is postulated on a society populated by different groups of agents with distinct cultural traits and technologies. Each time period a policy game is played between individual agents and a socio-economic policy maker (the government). Institutions represent the relative political power of different groups in civil society to affect policy decisions. Culture represents the distribution of values and preferences of these groups over policy decisions in society. The government's choice maximizes a social welfare function which encodes the distribution of political power between the societal groups (institutions) and their preferences and values (culture). A set of government policies and agents' actions arise as societal equilibrium outcomes.

Institutions evolve as the result of a process of optimal political delegation, changing the distribution of political power to internalize externalities, lack of commitment, and other distortions leading to an inefficient social equilibrium outcomes. As a consequence, residual decision rights over public policy tend to be delegated to those political groups which are better able (or have the highest incentives) to internalize the externalities affecting the policy game. Culture evolves over time, following socialization and cultural transmission processes whose incentives are in turn affected by equilibrium outcomes of the policy game played in society. The interdependence between institutions and culture is the fundamental factor determining their joint dynamics and their effects on economic activity, e.g., on long-run economic growth.

In such a setting, we characterize the cultural and institutional dynamics of the socio-economic system. From a normative perspective, we show that, even though institutional change is designed to respond to the inefficiencies of equilibrium outcomes, the societal equilibrium at the stationary state of the dynamics is not necessarily efficient and we characterize the determinants of the welfare properies of the dynamics. From a positive perspective, we characterize conditions under which cultural and institutional dynamics are complements, so that any e.g., institutional change which spurs economic activity is reinforced by the dynamics of culture and institutions; and conditions under which on the contrary culture and institutions are substitutes, hence weakening the effects of institutional change. We show how interesting examples of complex dynamics can emerge from the interaction between cultural and institutional change and how these qualitative dynamics depend on whether culture and institutions are complements or substitutes. These dynamical systems will generally tend to display forms of hysteresis of the equilibrium path as well as multiple stationary states and dependence on initial conditions (lack of ergodicity). We characterize a sufficient condition to rule out limit cycles and we show that i) local stability requires a bound on the strength of complementarity between culture and institutions, ii) but oscillatory convergent paths may only occur when culture and institutions interact as substitutes.

We define the *cultural multiplier*, as the ratio of the total effect of institutional change to its direct effect, that is, the counterfactual effect which would have occurred had the distribution of cultural traits in the population remained constant after the institutional change. Similarly, the *institutional multiplier*, is defined as the ratio of the total effect of cultural change to its direct effect. We show that these multipliers have the same sign, which depends on whether culture and institutions act as complements or substitutes, positive in the first case and negative in the second. The cultural and institutional multipliers are conceptual constructs defined to distinguish (and potentially measure) the relative contribution of culture and institutions to an economic outcome of interest, e.g., long-term economic growth, independently of the initial causal forcing variable. This analysis can complement, in important ways we identify and illustrate, the recent wave of causal analyses of either institutions or culture on future socio-economic prosperity, in *Historical economics* and *Persistence studies*.[2] Indeed, when the multipliers are large (positive or negative), the causal analysis of culture and institutions e.g., on economic activity, loses relevance to the study of the interactions between culture and institutions.

We specialize the abstract model of the joint evolution of culture and institutions into different models of the interaction of elites and civil society (or of the class struggle) as in the classic sociology and political sciences, notably after Marx (1867), Pareto (1901), Aron (1950a,b). Indeed, these models identify the distinct roles of culture, institutions, and of their interactions in different contexts of interests, elucidating several themes of great interest in the literature. In this paper we illustrate this studying the sustainability of *extractive institutions* and the formation of *civic capital* in two specific example societies.[3]

In the first example society, elites cultivate their preferences for leisure while extracting resources from the workers via taxation. We study the conditions under which the cultural and institutional dynamics maintain or reverse *extractive institutions*. We show that, in such society, depending on initial conditions, the elites might or might not have an interest in establishing less extractive institutions. When they do, they devolve part of the fiscal authority to workers, indirectly committing institutions to a lower tax rate. This in turn induces workers to exert an higher labor effort and hence to contribute more to fiscal authorities. We also show that, in this society, culture and institutions are complements: the devolution of fiscal authority to the

[2] See the *Handbook of Historical Economics*, Bisin and Federico (2021), for several surveys; especially, Valencia (2021), Cantoni and Yuchtman (2021), Voth (2021), and Bisin and Moro (2021).

[3] Other themes this model has been specialized to study include the protection of property rights (Bisin and Verdier, 2021), religious legitimacy (Bisin et al, 2021), cultural revivals (Iygun, Rubin, and Seror, 2021), and the industrialization process (Touré, 2021).

workers weakens the incentives of the elites to transmit their own cultural traits (their preferences for leisure); and smaller elites have higher incentives to devolve fiscal authority to workers.

In the second example society, a fraction of the population is endowed with a *civic capital* which has beneficial effects on the functioning of public governance structures, and hence on other members of the society. We study conditions under which the cultural and institutional dynamics favors or hinders the accumulation of civic capital in society. We show that, in this society, culture and institutions may act as substitutes: civic capital is more likely to spread when the degree of political representation is large and diffused; but the larger the diffusion of civic capital in society, the smaller the need to design institutional changes devolving formal power to prevent the mis-governance of public policies. In this context, an exogenous institutional change enlarging political representation (e.g., a form of democratization), may end up having its effects mitigated by regressive dynamics of the accumulation of civic capital .

The paper is structured as follows. In the next section we briefly discuss the related literature. In Section 2 we introduce, separately, the model of the dynamics of institutions and the model of the dynamics of culture. In Section 3 we study the joint dynamics of culture and institutions and we introduce the cultural and institutional multipliers as tools for the empirical analysis. Section 3 ends with a discussion of two important extensions, allowing for some forward looking behavior in the institutional design process and for (strategic) actions/policies driving cultural dynamics, e.g., the actions of a cultural leader. In Section 4, we introduce and study in some detailed example societies the dynamical interactions of elites and civil society. Finally, in Section 5 we conclude, drawing and discussing some empirical implications of our analysis, as a complement to the causal studies of the effects of culture and institutions in Historical Economics and Persistence studies.

1.1 Related literature

We model institutions as a representation of the relative power of different political groups. This is in line with the pathbreaking series of contributions by Acemoglu, Johnson, and Robinson and others,[4] but it also diverges from it in several ways. In Acemoglu (2003) and in Acemoglu and Robinson (2006), e.g., institutions are a representation of political pressure groups exercising the power to control social choice; and institutional change takes the form of voluntary transfer of power across groups, typically under threat of social conflict. In this paper instead, we depart from the notion of political power as concentrated in one single political group. We represent institutions as Pareto weights associated to the different groups in the social choice problem. This allows us to view institutional change as more incremental (formally, a continuous rather than

[4]See Acemoglu, Johnson, and Robinson (2006), Acemoglu, Egorov, and Sonin (2021) for surveys. See also Modica and Levine (2021) and Bowles et al (2021) for an alternative evolutionary approach based on external conflict.

a discrete change in political control) than just revolutions and regime changes;[5] in line, e.g., with the wealth of examples of institutional evolution through gradual and piecemeal changes in Mahoney and Thelen (2010).[6] It also allows us to enlarge the scope of our analysis from the study of transitions between autocracy and democracy and viceversa, which e.g., Acemoglu and Robinson (2000, 2001, 2006) concentrate on.

Our modeling of institutional change is also related to the cited contributions by Acemoglu, Johnson, and Robinson and others in that institutional change operates as a commitment mechanism. But our analysis focuses on mechanisms designed to internalize inefficient political choices rather than to limit the threat of social conflict.[7] The process of institutional change we study in this paper is also characterized by some form of myopia, to simplify the analysis.[8] Interestingly, however, myopic institutional change may be also factually motivated, e.g., in the historical process which underlies the emergence of democracy; see Treisman (2017).[9]

As far as culture is concerned, we conceptualize it as preference traits, norms, and attitudes and we allow for several social selection forces. In fact, the replicator dynamics we postulate can be micro-founded from i) evolutionary models using various payoff imitation protocols (Hofbauer, 1995; Helbing, 1992; Bjornerstedt and Weibull, 1996; Weibull, 2005); from ii) indirect evolutionary models of preference dynamics (Güth and Yaari, 1992; Güth, 1995; and, for applications to specific contexts, Alger and Weibull, 2013; Besley, 2017, 2020, Besley and Persson, 2019, 2020); and from evolutionary anthropology models of cultural transmission (Cavalli-Sforza and Feldman, 1973, 1981; Boyd and Richerson, 1985; and, for an economic approach with parental socialization choice, Bisin and Verdier, 1998, 2000a,b, 2001).[10]

A number of papers study theoretically the implications of the interactions between culture and institutions for economic activities. These papers however typically focus each on a distinct context-specific instance of these interactions, rather than on an abstract model of the political

[5]On the other hand, our analysis can be extended to account for (a smoothed formulation of) revolutions and regime changes; see Bisin and Verdier (2021).

[6]Relatedly, see Gradstein (2007, 2008) and Guimaraes and Sheedy (2016) who ground the study of institutions in the theory of coalition formation.

[7]For specific positions along these lines pertaining to the explanation of the extension of the franchise in early nineteenth century England, see Lizzeri and Persico (2004), Acemoglu and Robinson (2000, 2001, 2006) and Conley and Temimi (2001).

[8]A fully forward looking model of institutional change is analytically intractable when joined with cultural dynamics; though forward looking institutional change per se is studied by Lagunoff (2009) and Acemoglu, Egorov and Sonin (2015). In Section 4 we extend our model to accomodate some forward looking behavior to encompass "slippery slope" arguments.

[9]Specifically, Treisman (2021) argues that in the majority of the events he classifies democracy has been the outcome of miscalculation and lack of anticipation of the effects of the process set in motion by institutional change. Specifically, in several instances, the "incumbent initiates a partial reform, [...] but cannot stop" (see Table 2 in the paper), a representation which closely maps our modeling of myopic institutional change.

[10]See Bisin and Verdier (2011) and (2021) for surveys and discussions; and the Appendix for the formal derivations.

economy of elites and civil society as in the present paper; notably, e.g., work norms and the welfare state (Bisin and Verdier, 2000b), norms of cooperation and legal systems (Tabellini, 2008b), preference for patience and work ethics, and labor markets in the Industrial Revolution (Doepke and Zilibotti, 2008), trust and regulation (Aghion, Algan, Cahuc, and Shleifer, 2010), organizational culture and incentives (Besley and Ghatak, 2017; Besley and Persson 2020), civic culture and democratic institutions (Ticchi, Vindigni, and Verdier 2013; Besley and Persson, 2019), individualism and market organization (Davis and Williamson, 2016).[11]

2 The society

Consider a society with a continuum of agents separated into distinct groups defined in terms of relevant characteristics, i.e., political power and cultural traits, as well as, possibly, resources and technologies. We assume that political and cultural groups are aligned and indexed by $i \in I$.[12] We restrict for simplicity to dichotomous groups, $I = \{1, 2\}$, avoiding the issue of coalition formation in institutional design. The abstract choice of an agent belonging to group i is denoted a^i and the government's socio-economic policy choice is denoted p. Let $\mathbf{a} = \{a^1, a^2\}$ denote the vector profile of actions and let $\mathbf{e} = (\mathbf{a}, p)$.[13]

We characterize culture in this society in terms of the distribution of the population by cultural groups. Let q^i denote the fraction of agents of group i in the population, with $\sum_{i \in I} q^i = 1$. We adopt the shorthand $q^1 = q,\ q^2 = 1 - q$. Cultural groups are represented and characterized by direct preference traits, by norms and/or conventions they might abide by, ethnic and/or religious identities, and so on. Formally, this can be captured by the preferences of agents belonging to group i, written as an indirect utility function $u^i\left(a^i, p; \mathbf{a}, q\right)$, denoted compactly as $u^i(\mathbf{e}, q)$. The dependence of u^i on $\mathbf{a}$ captures indirectly any externality in the economy. The dependence of u^i on q captures instead indirectly the dependence of technologies and resources on the distribution of the population by groups. A natural example would have the externality being represented by the mean action in the population, $A = qa^1 + (1-q)a^2$, which furthermore showcases the possible interaction between the externality and the distribution of the population.

We conceptualize institutions as mechanisms through which social choices are delineated and

[11] See also Lindbeck (1995), Bidner and Francois (2011), Benabou, Ticchi, Vindigni (2015), Gorodnichenko and Roland (2021); and Alesina and Giuliano (2015), Acemoglu and Robinson (2021), Persson and Tabellini (2021), and Bisin and Verdier (2021) for surveys.

[12] In Section 4 we allow for groups which are distinct in terms of political power and cultural traits (with no substantial effects on the general analysis). With more than two groups the issue of coalition formation in institutional set-up and change becomes central.

[13] Both actions and policies are assumed to lie in a compact set. Of course policies might be multi-dimensional, an extension we avoid for simplicity. Also, without loss of generality we could add a parametrization of the component of economic institutions which acts directly on the economic environment. We avoid clogging the notation when not necessary.

implemented at equilibrium. Specifically, we model institutions as (Pareto) weights associated to the different groups in the social choice problem which determines policy making at equilibrium. Let $\beta^i \geq 0$, denote the weight associated to group i, with $\sum_{i \in I} \beta^i = 1$. Adopting again the shorthand $\beta^1 = \beta$, $\beta^2 = 1 - \beta$, the *social welfare function* used to evaluate public policies in social choice problem is:

$$W(\beta; \mathbf{e}, q) = \beta \, u^1(\mathbf{e}, q) + (1 - \beta) \, u^2(\mathbf{e}, q) \,. \tag{1}$$

Importantly, we do not interpret the social choice problem normatively, but rather as the indirect choice problem solved by the political process, where β encodes the relative power of the two groups.

2.1 Societal equilibrium and welfare

In this section we introduce the concept of equilibrium for our society, given institutions β and a distribution of the population by cultural group q. We then study the welfare properties of equilibrium.

At an equilibrium in society, each agent acts non-cooperatively with respect to all other agents and with respect to the economic policy choice. That is, given the policy choice p, $\mathbf{a}$ is a Nash equilibrium of the agents' choice problem. Economic policy is chosen to maximize the social welfare function, which encodes the relative power of the groups, without commitment: the policy maker cannot choose the policy p in advance of the choices of the economic agents. Formally, a *societal equilibrium* $\mathbf{e} = (\mathbf{a}, p)$ is a Nash equilibrium of the simultaneous game between agents and the policy maker, in an institutional set-up characterized by weights β and distribution by cultural group q:

$$\begin{array}{l} p \in \arg\max_p \; W(\beta; \mathbf{e}, q) \\ a^i \in \arg\max_a \; u^i(\mathbf{e}, q)\,, \quad i \in I = \{1, 2\} \end{array} . \tag{2}$$

A *social equilibrium* will generally not be efficient. Indeed, the agents' equilibrium choice $\mathbf{a}$ will not be efficient in general, given the policy choice p, because of the strategic interactions across agents and because of the externality directly embedded in the formulation of preferences. The policy choice p then adds a further layer of inefficiency, because of lack of commitment.[14] More precisely, a formal welfare analysis of societal equilibrium requires defining the *societal optimum* $\mathbf{e}^{eff}$,

$$\mathbf{e}^{eff} \in \arg\max \; W(\beta; \mathbf{e}, q)\,, \tag{3}$$

[14] See also Acemoglu (2003) and Belloc and Bowles (2017) for models of inefficient institutional dynamics in a different context.

and the equilibrium notion under commitment, the *societal commitment equilibrium*, $\mathbf{e}^{com}$, defined as the Stackelberg Nash equilibrium of the societal game,

$$\begin{array}{ll} \mathbf{e}^{com} & \in \arg\max \ W(\beta; \mathbf{e}, q) \\ & s.t. \quad a^i \in \arg\max_a \ u^i(\mathbf{e}, q), \quad i \in I = \{1, 2\} \end{array}. \tag{4}$$

Making the dependence on (β, q) explicit, the *societal equilibrium* and the *societal commitment equilibrium* can be denoted, respectively, by $\mathbf{e}(\beta, q) = \begin{pmatrix} \mathbf{a}(\beta, q) \\ p(\beta, q) \end{pmatrix}$ and $\mathbf{e}^{com}(\beta, q) = \begin{pmatrix} \mathbf{a}^{com}(\beta, q) \\ p^{com}(\beta, q) \end{pmatrix}$.

For regularity, we assume that utility functions are such that $\mathbf{e}(\beta_t, q_t)$, $\mathbf{e}^{com}(\beta_t, q_t)$ are continuous functions.[15] It is straightforward then to show that the *societal optimum*, the *societal equilibrium,* and the *societal commitment equilibrium* are generally distinct and weakly ranked in terms of welfare.

Proposition 1 *Given institutions β and a distribution of the population by cultural group q, the* societal equilibrium *is weakly inefficient, that is, weakly dominated by the* societal optimum*. Indeed, the* societal optimum *weakly dominates the* societal commitment equilibrium*, which in turn weakly dominates the* societal equilibrium.

This result is a straightforward consequence of the fact that, for any (β, q): i) problem (4), which defines a *societal commitment equilibrium*, is a constrained version of problem (3), which in turn defines a *societal optimum*; ii) any *societal equilibrium* satisfying (2) is always contained in the constrained feasible set of problem (4), which defines a *societal commitment equilibrium.*

To illustrate conditions determining whether the dominance relationships established in Proposition 1 are weak or strong, it is convenient to consider a simple society facing redistributive policies, in which the role of externalities and of lack of commitment is clearly apparent. We introduce this society in the Box in the following page and use it as a running example in this section.

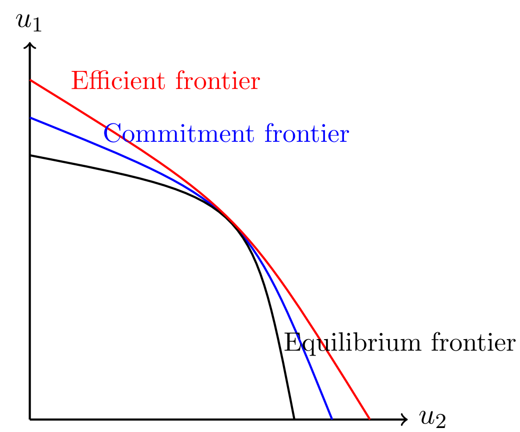
2.2 Institutional dynamics

We postulate institutions evolving over time as a mechanism to alleviate the inefficiency which plagues the *social equilibrium*, that is, the inefficiency due to the direct externality in the agents preferences and to the lack of commitment of the policy maker. In particular, while economic policies are chosen without commitment, society can commit to institutional change in the form of redistribution of political power across groups. Institutional change then occurs when redistributing power across groups leads to higher social welfare, evaluated with respect to the distribution

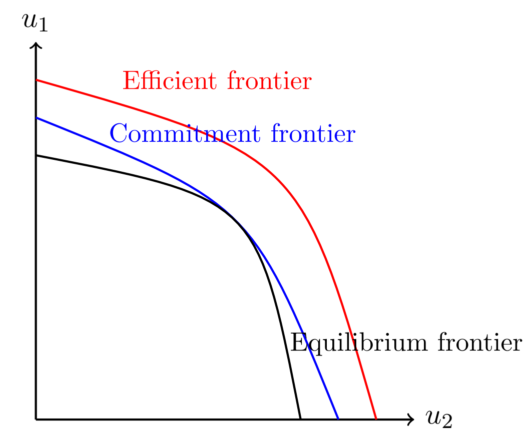
[15] See Appendix C for the obvious but stringent restrictions on fundamentals.

Box: A Redistribution Example - Welfare Analysis

Consider a society where fiscal policy is purely redistributive, taxes and subsidies are distortionary, and fiscal policy lacks commitment at a social equilibrium. The society is populated by two groups, $i \in (1, 2)$ of the same size. Each agents of group i is endowed with a production function h^i which maps his/her input a^i into output $h^i(a^i)$. Input of group $i = 1$ (resp. $i = 2$) is taxed linearly at rate p (resp. s); a negative rate is a subsidy. The government budget constraint is $pa^1 + sa^2 = 0$. The consumption of the two groups of agents - that is, their output net of taxes and subsidies - is: $h^1(a^1) - pa^1$; $h^2(a^2) - sa^2$, with no externality; and $h^1(a^1) - pa^1$; $h^2(a^2) + v(a^1) - sa^2$, with externality $v(a^1)$.



(a) Utility Frontiers: No Externality. Unique β^* such that $p(\beta^*, q) = 0$ and $\mathbf{e}(\beta^*, q) = \mathbf{e}^{eff}(\beta^*, q)$. Therefore, also $\mathbf{e}(\beta^*, q) = \mathbf{e}^{com}(\beta^*, q)$. For any $\beta \neq \beta^*$, $\mathbf{e}(\beta, q)$ is strictly dominated by $\mathbf{e}^{com}(\beta, q)$, which in turn is strictly dominated by $\mathbf{e}^{eff}(\beta, q)$.



(b) Utility Frontiers: Externality. $\mathbf{e}(\beta, q) \neq \mathbf{e}^{eff}(\beta, q)$, for any $\beta \in [0, 1]$. Unique β^E such that $\mathbf{e}(\beta, q) = \mathbf{e}^{com}(\beta, q)$. For any $\beta \neq \beta^E$, $\mathbf{e}(\beta, q)$ is strictly dominated by $\mathbf{e}^{com}(\beta, q)$, which in turn is strictly dominated by $\mathbf{e}^{eff}(\beta, q)$.

In the case in which the production choices of agents display no externalities - quadrant a), left - the *social equilibrium* is efficient if and only if there is no fiscal policy intervention. More precisely, in the society with no externality, there exist a unique set of institutions β^* such that no redistributive policy is implemented at a *social equilibrium* and the social equilibrium is efficient. Any other set of institutions $\beta \neq \beta^*$ induces a redistributive fiscal policy, which is inefficient due to the distortionary effects of taxes and subsidies and to lack of commitment. Furthermore, for any $\beta \neq \beta^*$ the *social equilibrium* is strictly dominated by the *commitment social equilibrium* which is itself inefficient, but only because of the distortions. In the case in which, on the contrary, the production choice of agents of one type has a positive externality on the income of agents of the other - quadrant b), right - *social equilibria* are always inefficient. There is a unique set of institutions β^E such that the *social equilibrium* coincides with the *commitment equilibrium*. At β^E, the social commitment equilibrium has the property that the marginal benefit of the externality is equal to the marginal cost induced by the distortionary subsidy. For any $\beta \neq \beta^E$ the *social equilibrium* is strictly dominated by the *commitment equilibrium*, and both are inefficient.

of power prior to the change. The mechanism driving the institutional dynamics of society, therefore, is akin to optimal political delegation: more political power is delegated to those groups which are better able (or have the highest incentives) to internalize the externalities affecting the policy game.

More precisely and operationally, turning to an index t for an explicit notation for time, a given current set of institutions in period t, β_t, induces the social preference order internalized by the policy maker at t. Future political and economic institutions, β_{t+1}, are designed at the end of period t, to maximize the current social welfare function by means of future policy choices, at $t+1$. Assuming that institutional design is myopic, that is, institutions are designed for the future as if they would never be designed anew in the forward future, institutions at time $t+1$ are designed at time t as a solution to:

$$\max_{\beta_{t+1}} W\left(\beta_t; \mathbf{e}(\beta_{t+1}, q_{t+1}); q_{t+1}\right). \tag{5}$$

In words, the societal equilibrium induced by institutions β_{t+1} at $t+1$ is chosen to maximize the social welfare induced by institutions β_t.

The formal characterization of the resulting institutional dynamics has a fundamental structure with a clear and straight interpretation. At any time t, current institutions β_t induce the policy choice $p(\beta_t, q_t)$ at equilibrium. But social welfare evaluated at weights β_t is highest under policy $p^{com}(\beta_t, q_t)$. Therefore, institutions at time $t+1$, β_{t+1}, are designed to induce a policy $p^{com}(\beta_t, q_{t+1})$. We assume, for regularity, that $p(\beta_t, q_t)$ is monotonic in β_t and indeed, without loss of generality, that it is increasing in β_t.[16] In this case, we show (in the Appendix) that, at the solution of problem (5), β_{t+1} is chosen so that:

$$p^{com}(\beta_t, q_{t+1}) = p(\beta_{t+1}, q_{t+1}), \tag{6}$$

unless Equation (6) does not have a solution, in which case β_{t+1} is chosen at a corner, either 0 or 1.

Equation (6) is the fundamental equation governing the (interior) dynamics of institutions in our analysis. For a given society characterized by institutions β and a distribution by cultural trait q, these dynamics of institutions can be characterized by studying the function $P(\beta, q) := p^{com}(\beta, q) - p(\beta, q)$, which is an indicator of the extent of the policy commitment problem faced by such society. More specifically, the absolute value of $P(\beta, q)$ indicates the intensity of the commitment problem as it reflects the distance between what can best be achieved under commitment and what is actually achieved at equilibrium. The sign of $P(\beta, q)$, on the other hand, indicates the direction of the institutional change in β needed to ameliorate the commitment problem.

[16] That is, we label group 1 (with weight β) the group preferring higher values of the policy p. See Appendix C for the obvious but stringent restrictions on fundamentals which guarantee that monotonicity is satisfied and for a generalization of the analysis relaxing it.

The properties of the dynamics of institutions β_t which are most relevant in our subsequent analysis are collected in the following:[17]

Proposition 2 *Given q, the dynamics of institutions have at least one stationary state. Any interior stationary state β^* obtains as a solution to $P(\beta, q) = 0$. The boundary stationary state $\beta = 1$ obtains when $P(\beta, q) \mid_{\beta=1} > 0$; while the boundary stationary state $\beta = 0$ obtains when $P(\beta, q) \mid_{\beta=0} < 0$.[18] In the continuous time limit, the dynamics satisfy the following properties:*

- *if $P(\beta, q) > 0$ for any $\beta \in [0, 1]$, then $\beta = 1$ is a globally stable stationary state;*
- *if $P(\beta, q) < 0$ for any $\beta \in [0, 1]$, then $\beta = 0$ is a globally stable stationary state;*
- *any boundary stationary state is always locally stable;*
- *any interior stationary state β^* is locally stable if $\frac{\partial P(\beta^*, q)}{\partial \beta} < 0$.*

We consider as an illustration the redistribution example introduced in the Box in previous section and continuing in the following page.

The characterization of the stationary states of the dynamics of institutions we obtained in Proposition 3 can be shown to imply that while the institutional dynamics do drive society towards efficiency, it is not generally the case that institutions are efficient in a stationary state, nor that institutions lead to Pareto improvements. Indeed, the dynamics of institutions has an efficient stationary state iff there exist β^* such that $\mathbf{e}(\beta^*, q) = \mathbf{e}^{eff}(\beta^*, q)$. If no such β^* exists, all stationary states are inefficient. Furthermore, the dynamics of institutions does not generally converge to a stationary state which Pareto dominates the initial institutional state of society, β_0. More specifically, if the dynamics of β_t converges to an interior stationary state β, then $e(\beta, q) = e^{com}(\beta, q)$. But in this case the commitment equilibrium utility frontier is (weakly) negatively sloped and the slope of the societal equilibrium frontier concides with it at β. As a consequence, the dynamics from any β_0 in an neighborhood of the stationary state drives the utilities of the two groups in opposite directions.

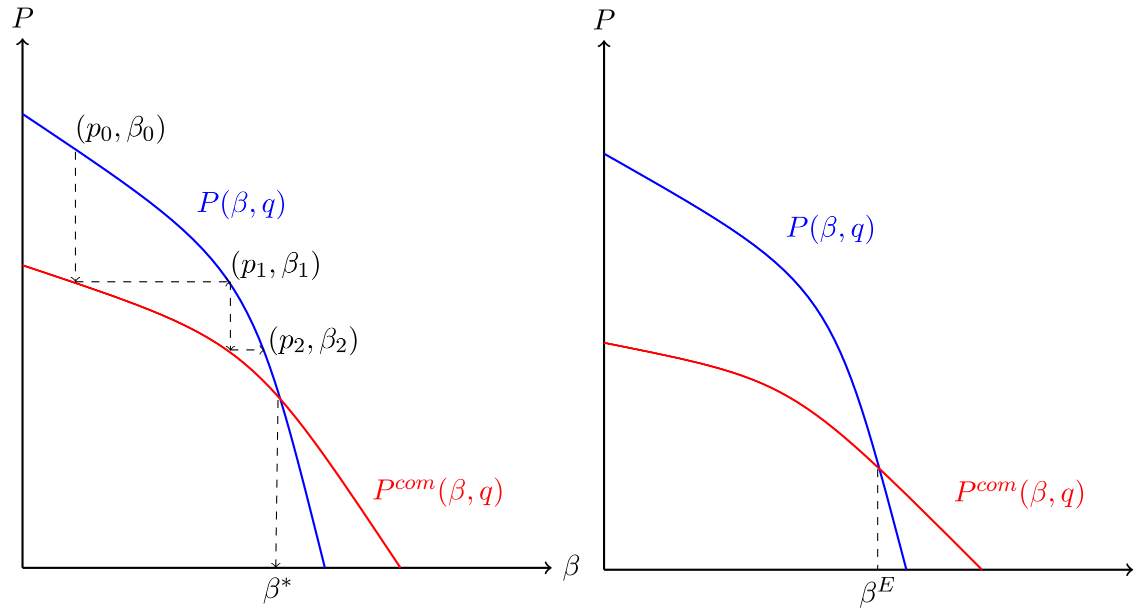
In the redistribution society with no externality, the dynamics converges to $e(\beta^*, q)$, which is efficient, but does not constitute a Pareto improvement for some initial institutional state β_0 close enough to β^*. In the case with the externality, the dynamics converges to $e(\beta^E, q)$, which is instead not efficient and, as in the case with no externality, does not contitute a Pareto improvement for some initial institutional state β_0 close enough to β^E.

[17] A more complete global stability analysis is not particularly complex but tedious. We relegate it to Appendix A, Proposition A.1.

[18] Note that we arbitrarily define $\beta = 1$ (resp. $\beta = 0$) as an *interior stationary state* if $P(\beta, q^i) \mid_{\beta=1} = 0$ (resp. $P(\beta, q^i) \mid_{\beta=0} = 0$).

Box: A Redistribution Example - Dynamics

Dynamics of Institutions: $\frac{\beta_{t+1}}{1-\beta_{t+1}} = \frac{\beta_t}{1-\beta_t} D(\mathbf{e}, q)$, with $P(\beta, q) \gtreqless 0$ if $D(\mathbf{e}, q) \gtreqless 1$.



(a) Dynamics of Institutions: No Externality. At $\beta = \beta^*$, $P(\beta, q) = 0$, $D(\mathbf{e}, q) = 1$, $\epsilon^1 + \epsilon^2 = 0$; where $\epsilon^1 = \mid \frac{p}{a^1} \frac{da^1}{dp} \mid_{p=p^{com}}$, $\epsilon^2 = \mid \frac{s}{a^2} \frac{da^2}{ds} \mid_{s=s^{com}}$.

(b) Dynamics of Institutions: Externality. At $\beta = \beta^E$, $P(\beta, q) = 0$, $D(\mathbf{e}, q) = 1$, $v'(a^{1,com}) = p$.

In this society, in both the cases with and without the externality, Equation 6 governs the dynamics of institutions. In the case in which there's no externality in production, $P(\beta, q) \gtreqless 0$ if $\beta < \beta^*$; while in the case with externality, this is the case if $\beta < \beta^E$. In the first case, a policy of redistribution introduces distortions which reduce the level of the taxed input and stimulate the level of the subsidized input. Institutional change is then induced in such a way as to provide more power to the group which is less distortionary at the margin. When $\beta_t < \beta^*$, the taxed group 1 is the group most affected by the redistributive distortions, in that the elasticity of this group input with respect to its tax, ϵ_1, is higher in absolute value than the elasticity of group 2 with respect to its subsidy, ϵ_2. Consequently more political decision rights are delegated to group 1, by increasing β_{t+1} with respect to β_t. Conversely when $\beta_t > \beta^*$, it is group 2 which is taxed and is most affected by the redistributive distortions. Consequently, Group 2 gets more political decision rights and β_{t+1} is decreased; see quadrant a), left. The case with externality is similar: increased political delegation is obtained by the group generating the positive externality as long as the marginal benefit of the externality is greater than the marginal cost induced by the distortionary subsidy. This is the case for $\beta < \beta^E$; see quadrant b), right. In the redistribution example, β_t converges to the unique interior stationary state, β^* and β^E, respectively, in the society without and with the externality.[a]

[a] See the onlline appendix for further details.

2.3 Cultural dynamics

We postulate a dynamics of the distribution of the population by cultural trait following a simple (logistic) replicator dynamics. This functional form is the formal representation of several interesting distinct cultural selection processes, as we noted when discussing the related literature in Section 1.1. Formally, given β_{t+1}, the dynamics of the distribution by cultural group q_t are governed by a difference equation of the following form:

$$q_{t+1} - q_t = q_t(1 - q_t)S(\beta_{t+1}, q_{t+1}); \tag{7}$$

where $S(\beta_{t+1}, q_{t+1})$ represents the relative strength of group $i = 1$ in terms of its ability to spread in the population. Reflecting an abstract social selection process on cultural traits or norms of behaviors, $S(\beta, q)$ depends on the societal equilibrium set of actions and policy $\mathbf{e}(\beta, q) = [\mathbf{a}(\beta, q); p(\beta, q)]$. Typically, $S(\beta, q)$ takes the form

$$S(\beta, q) = h^1(\mathbf{e}(\beta, q), q) - h^2(\mathbf{e}(\beta, q), q); \tag{8}$$

where $h^i(\mathbf{e}, q)$ is an appropriate *cultural fitness* function of trait i in the population. In a pairwise comparison random matching imitation context (Weibull 1995), for instance, $h^i(\mathbf{e}, q)$ is simply proportional to the utility $u^i(\mathbf{e}, q)$ of agents of type i. In an indirect evolutionary approach, $h^i(\mathbf{e}, q)$ represents the material fitness at the social equilibrium $\mathbf{e}(\beta, q)$ for agents of type i; that is, with preferences $u^i(\mathbf{e}, q)$. In the cultural transmission models by Bisin and Verdier (2000a),2000b), 2001), paternalistic parents of the two cultural types spend costly resources to bias the process of preference acquisition of their children. In this case, unpacking utility functions so that $u^i(\mathbf{e}, q) = u^i\left(a^i, p; \mathbf{a}, q\right)$, and denoting $\Delta V^i(\beta, q) = u^i(a^i, p; \mathbf{a}, q) - u^i(a^j, p; \mathbf{a}, q)$, one gets

$$h^1(\mathbf{e}, q) = w\left[\Delta V^1(\beta, q) \cdot (1 - q)\right] \quad \text{and } h^2(\mathbf{e}, q) = w\left[\Delta V^2(\beta, q) \cdot q\right] \tag{9}$$

with $w(.)$ an increasing function. In words, the *cultural fitness* of trait i is increasing in the socialization gain of parents of trait i.

We assume for regularity that $S(\beta, q)$ is a continuous function. Generally, and independently of the underlying specific cultural selection process, we can then characterize the dynamics of the distribution by cultural group q_t as follows.[19]

Proposition 3 *The dynamics of the distribution of culture q_t have at least the two boundaries as stationary states, $q = 0$ and $q = 1$. Any* interior stationary state $0 < q^* < 1$ *obtains as a solution to $S^i(\beta, q) = 0$. In the continuous time limit, the dynamics satisfy the following properties:*

- *if $S(\beta, q) > 0$ for any $q \in [0, 1]$, then q_t converges to $q = 1$ from any initial condition $q_0 > 0$;*

[19] We collect here the properties of the dynamics of culture which are most relevant in our subsequent analysis. We relegate a more complete analysis to Appendix A, Proposition A.2.

- *if* $S(\beta, q) < 0$ *for any* $q \in [0, 1]$*, then* q_t *converges to* $q = 0$ *from any initial condition* $q_0 < 1$*;*

- *if* $S(\beta, 1) > 0$*, then* $q = 1$ *is locally stable ;*

- *if* $S(\beta, 0) < 0$*, then* $q = 0$ *is locally stable;*

- *any interior stationary state* $q^* (\beta)$ *is locally stable if* $\frac{\partial S(\beta, q^*)}{\partial q} < 0$*.*

When $S(\beta, q)$ has the form in (8), any interior steady state $q^* (\beta)$ satisfies $h^1 (\mathbf{e}(\beta, q^*), q^*) = h^2 (\mathbf{e}(\beta, q^*), q^*)$. When furthermore (9) is satisfied, any interior cultural stationary state $q^* (\beta)$ is obtained as a solution to:[20]

$$\frac{\Delta V^1(\beta, q)}{\Delta V^2(\beta, q)} = \frac{q}{1-q}. \tag{10}$$

3 Joint evolution of culture and institutions

In this section we study the dynamics of culture and institutions in the society introduced in the previous section, highlighting conditions under which interesting qualitative dynamical paths, like dependence on initial conditions, limit cycles, or other oscillatory dynamics, may or may not arise. We then introduce the cultural and institutional multipliers as useful tools for the analysis of these dynamics. We draw implications for the study of the effects of culture and institutions on economic variables of interest, as a complementary tool to the causal methods largely adopted in *Historical economics* and, particularly, in *Persistence studies*. Finally, we discuss some relevant extensions.

3.1 Dynamics

The model of cultural and institutional change introduced in the previous section delivers dynamics governed by the system of m difference equations (6,7). More precisely, once equation (6) is modified to allow for corners (see the Appendix), the dynamical system becomes:

$$\beta_{t+1} = \begin{cases} \beta \text{ such that } p^{com}(\beta_t, q_{t+1}) = p(\beta, q_{t+1}) & \text{if it exists,} \\ \left[\begin{array}{ll} 1 & \text{if } p^{com}(\beta_t, q_{t+1}) > p(\beta, q_{t+1}),\ \forall 0 \leq \beta \leq 1 \\ 0 & \text{if } p^{com}(\beta_t, q_{t+1}) < p(\beta, q_{t+1}),\ \forall 0 \leq \beta \leq 1 \end{array} \right. & \text{else} \end{cases} \tag{6'}$$

$$q_{t+1} - q_t = q_t(1 - q_t)S(\beta_{t+1}, q_{t+1}); \tag{7}$$

[20] In the Appendix, we show that, for every value of β, the cultural dynamics are monotonic and converge towards a unique interior stationary state $q^*(\beta)$, under strong enough "cultural substitution" (Bisin and Verdier, 2001); that is, when minority groups have higher marginal incentives to engage in socialization than majority groups in society.

Any *interior* stationary state of the system (6',7), (β^*, q^*), solves:

$$P(\beta, q) = S(\beta, q) = 0. \tag{11}$$

While the explanatory power of this system is best manifested once it is specialized to the study of phase-diagrams in specific societies, in Section 4,[21] a series of interesting results can be obtained even at this level of generality.

With regard to stationary states,while it is not always the case that an interior stationary state exists, we can show that

Proposition 4 *The dynamical system (6',7) has at least one stationary state.*

Clearly, multiple stationary states are possible in the general non-linear system. In Section 4 we will study interesting examples of non-ergodic dynamics, where different stationary states are reached by different basins of attractions in the space of initial conditions.

With regards to local stability, the formal characterization of the (possibly complex) dynamics of a two-dimensional system are of course well understood. We concentrate on identifying conditions which have a clear interesting interpretations in terms of the properties of cultural and institutional change. To this end we assume for regularity the following separability condition: $u^i\left(a^i, p; a, q^i\right) = v^i\left(a^i, p\right) + H^i(p; a, q^i)$ in the following analysis.

First of all,

Proposition 5 *The condition,*

$$\frac{\partial P(\beta^*, q^*)}{\partial \beta}, \frac{\partial S(\beta^*, q^*)}{\partial q} < 0, \tag{12}$$

is sufficient to guarantee no limit cycles in the neighborhood of an interior stationary state (β^*, q^*) *of the dynamical system (6',7).*

Condition (12) is an implication of the Bendixon Negative Criterion and has a clear intuitive interpretation. It requires that, in a neighborhood of the stationary state, institutional change has social decreasing returns as a mechanism to internalize the externalities of the equilibrium, $\frac{\partial P(\beta^*,q^*)}{\partial \beta} < 0$, and similarly that the selective forces driving cultural change also have decreasing returns, $\frac{\partial S(\beta^*,q^*)}{\partial q} < 0$, hence favoring cultural diversity in a neighborhood of q^*.[22]

Condition (12) guarantees a substantial reduction in the complexity of the dynamics of the system but is not sufficient for local stability, which instead also requires conditions involving the complementarity or substitutability of culture and institutions, in a precise sense which we

[21] See Bisin and Verdier (2021) for a discussion of the role of phase diagrams in Historical Economics and several examples.

[22] This is the case, for instance, under "cultural substitution" in cultural transmission models; see Footnote 20.

define next.[23] Let $\beta = \beta(q)$ be the stationary state manifold associated with the equation in (6'); that is, such that $P(\beta(q), q) = 0$. Let $q(\beta)$ be the stationary state manifold associated with equation (7); that is, such that $S(\beta, q(\beta)) = 0$. (Clearly, any interior stationary state lies at the intersection of the manifolds $\beta(q)$ and $q(\beta)$.) We then say that,

Institutional and cultural dynamics are locally complementary at an interior stationary state (β^*, q^*), when the stationary state manifolds $\beta(q)$ and $q(\beta)$ of the dynamical system (6',7) have slopes of the same sign at (β^*, q^*):

$$sign\left(\frac{d\beta(q^*)}{dq}\right) = sign\left(\frac{dq(\beta^*)}{d\beta}\right); \tag{13}$$

conversely they are locally substitutes at (β^*, q^*) when the slopes have opposite signs.[24]

Indeed, $\frac{d\beta(q)}{dq}$ represents the relationship between group 1's size and its ability to internalize the externality at equilibrium when social welfare favors the group's policy preferences; while $\frac{dq(\beta)}{d\beta}$ represents the relationship between the group's weight in social welfare and the strength of its cultural diffusion. When, by way of illustration, these two terms are both positive at a stationary state (the negative case is symmetric), institutional and cultural dynamics are locally complementary in the sense that the political power of a group breeds into its size which in turn breeds into political power, in a dynamic feedback.[25]

We now show how complementarity and substitutability turn out to be determinant factors governing the dynamics of culture and institutions and the local stability of system system (6',7),

Proposition 6 *Suppose condition 12 is satisfied. Then local substitutability of the institutional and cultural dynamics at an interior steady state (β^*, q^*) of the dynamical system (6',7) is sufficient for local stability. Under local complementarity, instead, local stability obtains if*

$$\frac{d\beta(q^*)}{dq}\frac{dq(\beta^*)}{d\beta} < 1. \tag{14}$$

Furthermore, in this case, the local dynamics show no converging oscillatory dynamics.

[23] The extension of this analysis to corner stationary states is notationally involved in the general case, with minor additional insights.

[24] All these conditions are useful in applications, as we will show in Section 4, in that they are generally easy to check.

[25] When interior cultural stationary states $q^*(\beta)$ are obtained as a solution to (10), that is, by cultural transmission as in Bisin and Verdier (2000a, 2000b, 2001), the complementarity condition on the slopes of $\beta(q)$ and $q(\beta)$ can be shown to require that $\frac{d\frac{\Delta V^1(p)}{\Delta V^2(p)}}{dp}$ and $\frac{\partial P(\beta^*, q^*)}{\partial q}$ have the same sign (see Appendix A for details). When, e.g., these two terms are both positive, i) the first guarantees that an increase in the institutional weight of the group which better internalizes the externality increases the share of this group in the population by increasing the incentives to cultural transmission; ii) the second guarantees that a larger fraction of the group in the population induces institutional change leading to a higher weight of this same group.

When both institutional and cultural change display decreasing returns at the stationary state - when Condition (12) is satisfied - substitutability implies that culture and institutions dampen each other on a convergence path, thereby guaranteeing local stability, though possibly inducing oscillatory dynamics. Under complementarity, instead, culture and institutions reinforce each other's dynamics. A strong complementarity might therefore have amplification effects, inducing dynamics which diverge towards the corners of the dynamical system: Condition (14) is effectively a bound on the strength of the complementarity of culture and institutions, which is required for stability. The (monotonicity of) reinforcement effects of complementary institutional and cultural dynamics, however, rule out oscillatory dynamics along a convergence path.

3.2 The cultural and institutional multipliers

A thriving literature in economics and political science studies which factors may account for various long-run variables of interest; per-capita income, first of all, but also public goods provision, political participation, and several other measures of socio-economic prosperity. Institutions and culture are often considered possible explanatory factors and historical natural experiments and/or instrumental variables are used to document the causal effects of either one or the other on the relevant socio-economic variable of interest.[26] In the context of these *persistence studies* disregarding the interactions between culture and institutions can be problematic, e.g., when the same institutional change may have differential effects according to different cultural environments. Most importantly, a better understanding the factors determining the long-run variables of interest requires an empirical assessment of the relative contributions of culture and institutions to their dynamics. Indeed, even when the causal analysis is valid, the relative contribution of culture and institutions is generally independent of its identified cause.

To this end, in this section, we introduce the concepts of of *cultural* (resp. *institutional*) *multiplier*, which we will then exploit when we specialize our analysis to the study of phase-diagrams in specific societies, in Section 4.[27] We define the *cultural* (resp. *institutional*) *multiplier* as the ratio of the long-run change in institutions (resp. culture) relative to the counterfactual long-run change that would have happened had the cultural composition (resp. institutional set-up) of society remained fixed. For ease of exposition, we shall concentrate on the *cultural multiplier*, under the understanding that symmetric arguments and conditions hold for the *institutional multiplier*. In fact, an implication of the following analysis is that the *cultural* and the *institutional* multiplier have the same sign, which depends on the complementarity/substitutability of culture

[26] See, most notably, Acemoglu, Johnson and Robinson (2001), Diamond (2005), Acemoglu and Robinson (2010), Dell (2010), for institutions; and Tabellini, (2008a), Nunn and Wantchekon (2011), Chen, Wang, Yan (2014) for culture. See the references in footnote 2 for surveys, some of them critical, and Przeworski (2004), Abad and Maurer (2021).

[27] See Bisin and Moro (2021), Cantoni and Yuchtman (2021), and Voth (2021) for a detailed discussion of these methodological issues.

and institutions.

Consider a long-run socio-economic variable which depends indirectly (in reduced form) from culture and institutions, $A(\beta, q)$. In fact, to illustrate the analysis, consider the case in which both β and q have a positive effect on A.[28] We distinguish in turn two comparative dynamics exercises: the first where a change $d\gamma$ in a parameter perturbs locally an interior stable stationary state (β^*, q^*), inducing a dynamical convergence path in culture and institutions and hence in A; the second one where we follow the dynamics of culture and institutions from an initial condition (β_0, q_0) in the basin of attraction of a stationary state (β^*, q^*), and hence from $A(\beta_0, q_0)$ to $A(\beta^*, q^*)$.

Local change $d\gamma$. Adding explicit reference to γ in the notation, we normalize the arbitrary components of the comparative dynamics environment we study so that a positive change in γ induces a process of convergence to a new steady state characterized by a larger *societal equilibrium* policy p (through a larger β).[29] In this context, the change in the long-run variable of interest A has two interrelated components $\frac{dA}{d\gamma} = \frac{\partial A}{\partial \beta}\frac{d\beta}{d\gamma} + \frac{\partial A}{\partial q}\frac{dq}{d\gamma}$, where $\frac{d\beta}{d\gamma} = \frac{\partial \beta}{\partial \gamma} + \frac{\partial \beta}{\partial q}\frac{dq}{d\gamma}$ and $\frac{dq}{d\gamma} = \frac{\partial q}{\partial \gamma} + \frac{\partial q}{\partial \beta}\frac{d\beta}{d\gamma}$ are the institutional and cultural components, respectively. Then,

The cultural multiplier *on institutional change, at a locally stable interior steady state* (β^*, q^*) m_{SS}*, is*

$$m_{SS} = \left(\frac{d\beta^*}{d\gamma}\right) / \left(\frac{\partial \beta^*}{\partial \gamma}\right) - 1 = \frac{\partial \beta}{\partial q}\frac{dq}{d\gamma}. \tag{15}$$

It follows that whether the multiplier is indeed positive crucially depends on culture and institutions being complements or substitutes.

Proposition 7 *Local complementarity (resp. substitutability) of the institutional and cultural dynamics at an interior stable stationary state* (β^*, q^*) *is sufficient for the cultural multiplier* m_{SS} *at* (β^*, q^*) *to be positive (resp. negative).*

The resulting cultural multiplier on the long-run variable A is $\frac{\partial A}{\partial q}\frac{dq}{d\gamma}$, which is positive when m_{SS} is positive. In the complementarity case, under our normalization, an increase in γ is set to induce an increase in β which, because of complementarity, is reinforced by an increase in q, inducing positive feedback dynamics. Any exogenous institutional change, through an increase in γ, is amplified by the associated cultural dynamics that interact with institutions, leading to $m_{SS} > 0$. Conversely, institutional changes would be hindered by cultural changes (i.e., the *cultural multiplier* m_{SS} is negative) when culture and institution are substitutes, that is, when the slopes of $\beta(q)$ and $q(p)$ have opposite signs.

[28]See Appendix B for the complete analysis, when A depends on agents' actions $\mathbf{a}$ and government policies p.

[29]Formally, we sign of the effects of a $d\gamma > 0$ so that it increases, locally at the steady state, both the policy p as well as the extent of the social externality or commitment problem: $\frac{dp^{com}(\beta^*, q^*; \gamma)}{d\gamma} > \frac{dp(\beta^*, q^*; \gamma)}{d\gamma} > 0$; Also, without loss of generality, we let members of group 1 (with institutional power β) aim at a relatively larger policy level, p: $\frac{\partial p(\beta^*, q^*; \gamma)}{\partial \beta} > 0$.

Global dynamics from (β_0, q_0). Consider an initial condition (β_0, q_0) in the basin of attraction of a stationary state (β^*, q^*). In this case the full dynamics of culture and institutions from (β_0, q_0) converges, by construction, to (β^*, q^*); that is, in particular, institutions converge to $\beta^* = \beta(q^*)$ and the long-run variable to $A(\beta^*, q^*)$. In the counterfactual case in which the cultural composition of society had remained fixed, the dynamics of institutions would have converged to $\beta(q_0)$ and A to $A(\beta(q_0), q_0)$. In such a case we say that,
The cultural multiplier on institutional change m_{DD} *from initial condition* (β_0, q_0) *in the basin of attraction of a stationary state* (β^*, q^*)*, is*

$$m_{DD} = \frac{\beta(q^*)}{\beta(q_0)} - 1. \tag{16}$$

The formal analysis of m_{DD} requires distinguishing between the *local complementarity* just defined and a more stringent *global complementarity*, defined as follows:
Institutional and cultural dynamics are globally complementary (resp. substitutes) when the steady state manifolds $\beta(q)$ *and* $q(\beta)$ *have slopes of the same sign (resp. opposite signs) for all values* $(\beta, q) \in [0, 1]^2$.
The next result shows that, even in the context of global analysis, whether the multiplier is indeed positive crucially depends on culture and institutions being complements or substitutes.

Proposition 8 *Under global complementarity of the institutional and cultural dynamics, the cultural multiplier* m_{DD} *from initial condition* (β_0, q_0)*, in the basin of attraction of* (β^*, q^*) *has the same sign as* $(q^* - q_0) \cdot \frac{d\beta(q)}{dq}$.

The resulting cultural multiplier on the long-run variable A is $\frac{A(\beta(q^*,q^*)}{A(\beta(q_0,q_0)} - 1$, which also is positive when m_{DD} is positive. As an illustration, suppose that culture and institutions are complements in the sense that $\frac{d\beta(q)}{dq}$ and $\frac{dq(\beta)}{d\beta} > 0$. Consider the process of convergence along a transition path from (β_0, q_0) to (β^*, q^*) with say, $q_0 < q^*$. Because of global complementarity between institutions and culture, one cannot have dampened oscillations in the basin of attraction of (β^*, q^*) and q_t increases monotonically from q_0 towards q^*. Along that transition path, this involves changes in β and q which reinforce each other: an increase in β induces a further increase in q which in turn feedbacks positively on the institutional weight β. Upon convergence, then, $\beta^* > \beta(q_0)$, where $\beta(q_0)$ is the counterfactual institutional steady state with culture fixed at q_0. This process implies therefore a positive institutional (and cultural) multiplier m_{DD} from the initial condition (β_0, q_0).

The cultural multipliers m_{DD} and m_{SS} illustrate different perspectives of the interactions between institutions and culture. The cultural multiplier m_{DD} encapsulates properties along the transition path of the joint dynamics between institutions and culture for a given society over time. Conversely, the cultural multiplier m_{SS} exhibits the steady state interactive effects between institutions and culture related to exogenous variations of some parameter γ. It emphasizes therefore the joint long-run effects of culture and institutions that may be observed across societies. In both cases, a positive *cultural multiplier* indicates the reinforcing effect of a change in institutions at a socio-economic equilibrium due to the endogeneity of culture.[30]

3.3 Extensions

In this section we discuss two important extensions of our model, allowing for i) some forward looking behavior in the institutional design process; and for ii) (strategic) actions/policies driving cultural dynamics, e.g., the actions of a cultural leader.

3.3.1 Forward looking institutional change

Institutional design is myopic in our model; that is, institutional change at time t, from β_t to β_{t+1}, is predicated under the assumption that institutions β_{t+1} will not change in the future. In other words, the mechanism driving institutional change does not anticipate the actual dynamics represented by Equation 6, from β_{t+1} to β_{t+2}, β_{t+3}, ... In particular, it could very well be that, with respect to the social welfare order associated to institutions β_t, these dynamics will lead to dominated outcomes. An institutional change mechanism with better forward looking ability might prevent or mitigate the logic of this institutional *slippery slope*, slowing down the pace of change or even stopping it altogether.

Our model can be extended in this direction, to account for some forward-looking institutional change. Fixing the cultural profile q prevailing in society, consider as an illustration a 1-step forward looking behavior: given institutions β_t in period t, institutions β_{t+1} are chosen to maximize the implied social welfare ordering at t, anticipating institutions β_{t+2} induced by Equation 6.[31] Under forward looking institutional change, institutions β_t balance the policy commitment gains from strategic delegation to β_{t+1} against its costs; but differently than in the myopic case, it also takes into account the benefits and costs of delegation from β_{t+1} to β_{t+2}. In this context, with strong enough regularity conditions to guarantee convexity of the institutional change problem, we show in Appendix B that an interior stable institutional steady state β^* under the myopic institutional dynamics in (the continuous time limit of) Equation 6 is also a steady state of the "1-step" forward-looking institutional dynamics. More interestingly, the 1-step forward looking institutional dynamics change converges to that steady state β^* but at a reduced speed, to mitigate the costs of the *slippery slope* in the dynamics of institutions. When the convexity conditions

[30] Another interesting class of comparative dynamics exercises consists of a change in a parameter γ which induces a change in the basin of attraction of two distinct stationary states of the system. While general results are difficult to obtain in these cases, in Section 5 we discuss the empirical implications of these class of exercises for the casual analysis of Average Treatment Effects; see also Bisin and Moro (2021).

[31] Of course $K-$step forward mechanisms can be characterized similarly.

are not satisfied, on the other hand, institutional change might not necessarily imply marginal institutional adjustments and the forward looking dynamics may get stuck at a point which remains far away from the myopic institutional steady state β^*.[32]

3.3.2 Centralized action/policies for cultural change

In our model, cultural dynamics are the results of evolutionary selective forces emerging from fully decentralized mechanisms. This perspective does not take into account the fact that cultural change can itself be influenced by centralized public institutions, such as states, churches, clans and community leaders. A recent literature has considered the role of these centralized cultural transmission agents in different settings.[33] An extension of our analysis along these lines would allow for the choice of some form of group level socialization effort, partly internalizing the effect of such effort on the dynamics of cultural change. In such a case, the joint dynamics of culture and institutions may result as the outcome of a dynamic game between public policy institutions, determining strategically the evolution of institutions β_t, and group-level cultural institutions, determining instead the evolution of cultural traits q_t.

While a full extension along these lines is analytically intractable, insights from the aforementioned literature on cultural leaders and comparative statics in games suggest two implications. First, the joint evolution of culture and institutions will have important forward looking dimensions along both the institutional and the cultural dynamics, creating important sources of multiple dynamic paths and non-ergodicity. Second, one may also expect the existence of additional positive (resp. negative) cultural and institutional multiplier effects when culture and institutions are strategic complements (resp. substitutes).

4 Political economy of elites and civil society

In this section we work out several specific societies, simple but rich enough to display some interesting cultural and institutional dynamics.[34] Other interesting societies, studied in terms of their cultural and institutional dynamics along the lines of this paper, are introduced in Bisin and Verdier (2021) and Bisin, Rubin, Seror, and Verdier (2021).

[32]See Acemoglu, Egorov and Sonin (2015) for a related analysis but in a context characterized by a discrete, hence not convex, set of possible institutions.

[33]See, for instance, Acemoglu and Jackson (2016), Almagro and Andres-Cerezo (2020), Carvalho (2016), Hauk and Mueller (2015), Prummer and Siedlarek (2017), Verdier and Zenou (2015, 2018).

[34]In all the examples we impose and exploit various regularity conditions without explicit mentioning them. Also, in some instances, we need to extend our analysis to consider a society in which political and cultural groups are distinct, but the extra notation and analysis will be straightforward. We refer to the Appendix for assumptions, notation, and analytical details.

4.1 Elites, workers, and extractive institutions

In this example we study *extractive institutions*, that is, a society where the power of a political group is exercised by extracting resources from the other group, e.g., via taxation. We study in particular conditions under which the cultural and institutional dynamics in this society maintain or reverse its extractive character. We are motivated in this analysis by the extensive work, spurred by Acemoglu and Robinson (2001), which identifies extractive institutions, as opposed to inclusive institutions, as one of the major obstacles which has hindered economic growth, and ultimately prosperity, across history.[35]

Consider a society populated by two groups, members of the elite and workers, with distinct cultural traits and technologies. In particular, the preferences of the members of the elite are shaped by cultural norms which let them value leisure greatly, more than workers. Hence in equilibrium, while both the elite and workers are endowed with the same technology which transforms labor into private consumption goods, workers will work and the elite will generally eschew labor and constitute a *leisure class*. In this society, taxes on labor income finance public good consumption, which is valued by both groups. But since members of the elite do not work, (fiscal) institutions are extractive in that only workers bear the fiscal weight of the public good.

Institutions lack commitment; that is, fiscal authorities choose the tax rate without internalizing its effect on workers' labor effort. This gives institutions generally an incentive to tax labor excessively. As a consequence, in this society, the elite might have an interest in establishing less-extractive institutions, by delegating part of the fiscal authority to workers. This would indirectly commit institutions to a lower tax rate, in turn inducing workers to exert an higher labor effort, contributing to higher income and public good production. Delegating fiscal authority to the workers however weakens the incentives of the elite to transmit its own culture and hence reduces the size of the leisure class; and in turn a smaller leisure class augments the incentives of the elite to delegate fiscal authority to workers. This form of complementarity drives the equilibrium dynamics of culture and institutions in this society.

Formally, let workers be group $i = 1$ and the elite be $i = 2$. Both groups can transform labor one-for-one into private consumption goods. Let a^i denote labor exerted by any member of group i. Let s denote the initial resources all agents are endowed with. Let p denote the (income) tax rate and G the public good provided by fiscal institutions. Preferences for group i are represented by the following utility function:

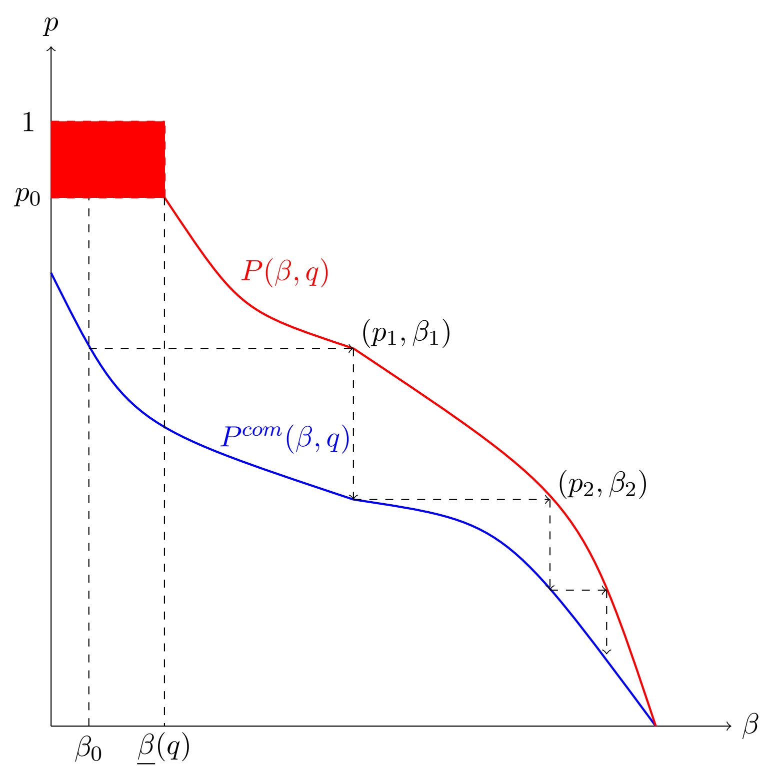
[35]See for example Engerman and Sokoloff (2002), Lange, Mahoney and vom Hau (2006), and Ogilvie (2021) for respectively discussions of extractive and inclusive institutions in Economic History. Ogilvie and Carus (2014) provides a good survey of the literature.

For analyses on specific regions, see for instance Huillery (2009), Nunn (2007, 2008), Michalopoulos and Papaioannou (2013, 2016), Tadei (2018) for Africa; Dell (2010), Arroyo-Abad and Luiten van Zanden (2016), Paredes-Fuentes (2016) for Latin America, Iyer (2010), Bogart and Chaudhary (2019) for South Asia, Dell and Olken (2020) for Indonesia.

$$u^i\left(a^i, G, p\right) = u\left(a^i(1-p)+s\right) + \theta^i v(1-a^i) + \Omega \cdot G.$$

To simplify the analysis, consider extreme preferences for leisure of the elite, $\theta^2 > \frac{u'(s)}{v'(1)} > 1 = \theta^1$. In this case, members of the elite never work, $a^2 = 0$, and consume their resources, s. Workers instead exert some effort level $a^1 > 0$ and consume in fact $a^1 + s$ units of the private consumption good. Both groups consume the public good G, in an amount equal the tax burden, to balance the budget of the fiscal institutions: $G = p\left[a^1 q + a^2(1-q)\right]$ where q is again the fraction of workers type $i = 1$.

Figure 3: Elites, workers and extractive institutions: Equilibrium policies and institutional dynamics



The *social equilibrium* and the *social commitment equilibrium* are then easily characterized, for any institutional set-up β,and distribution of the society by cultural traits q, as in Figure 3. Consider first the *societal equilibrium.* For β small enough, $\beta \leq \underline{\beta}(q)$, all policies p inducing no labor effort are a *societal equilibrium.* In this case, workers have so little power that the natural ex post incentive is to tax them to the extent that they do not provide any labor supply, $p \geq p_0$. For $\beta > \underline{\beta}(q)$ the *societal equilibrium policy* $p(\beta, q)$ is a decreasing function of β. The ex-post incentives to finance the public good through labor income taxes are lower when the workers' interest are better represented. At $\beta = 1$, labor income is taxed only inasmuch as it is necessary to finance the amount of public good preferred by workers themselves, say p^*. At the *social commitment equilibrium*, taxes are also declining with the political power of the workers β and, most importantly, they are lower than at the *societal equilibrium*, when the policy maker does not internalize the negative distortion of taxation on the tax base:

$$p^{com}(\beta, q) < p(\beta, q), \ \forall \beta < 1.$$

The institutional dynamics tend to internalize the inefficiency of the *societal equilibrium* which is due to lack of commitment, for any given cultural distribution in the population q. In this society, therefore, fixing $0 < q < 1$, for any $\beta < 1$, the institutional dynamics tend towards increasing the fiscal authority of workers, that is towards increasing β. This leads to reducing the excessive (and inefficient) tax rate p until it is optimal for the workers to do so. All fiscal authority ends up effectively in the hands of workers. The stationary state of the dynamics of institutions is therefore efficient. It does not account for the preferences of the elites, however, and hence it reduces their welfare with respect to a robust (large) set of initial institutional states of society, β_0; thereby illustrating the general result in Proposition 1.

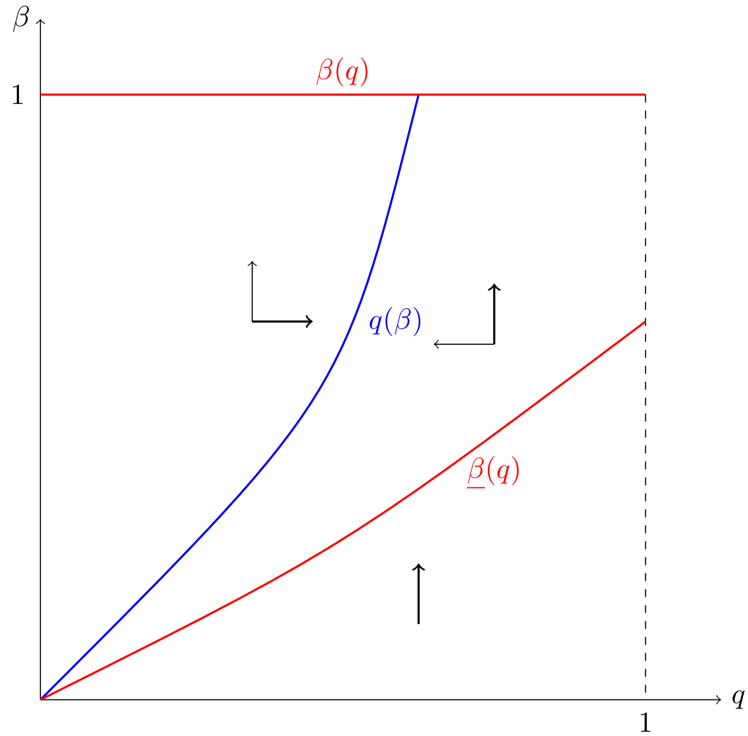
Given the long-run dynastic dynamics we are considering in this example, we adopt the intergenerational cultural transmission mechanism of Bisin and Verdier (2001). In this case, for every value of β the cultural dynamics tend to an interior stationary state $q(\beta)$, whereby q increases when $q < q(\beta)$ and decreases instead when $q > q(\beta)$. In other words, given the institutional set-up, both group tend to engage in more intense cultural transmission when their trait is relatively minoritarian in society. Furthermore, the relative incentives to socialization $\Delta V^1(p)/\Delta V^2(p)$ are decreasing in p. Indeed, as taxation leads to increased rent extraction on labor, leisure class norms are more likely to be transmitted than those of the workers: the larger the rents of the elites, the larger their socialization advantage.

The joint cultural and institutional dynamics of this society are concisely represented in the phase diagram in Figure 4. The curve along which β is constant is simply $\beta(q) = 1$: independently of the fraction of the workers' trait in the distribution, the dynamics of institutions has β_t increasing until all the power is to the workers, $\beta = 1$.

The curve along which q is constant, $q(\beta)$, is weakly increasing in β: more fiscal authority to workers leads to a lower equilibrium tax on labor p and therefore to an increase in the prevalence of the workers' cultural trait in the population in the long run dynamics of culture. Indeed, the lower is the tax rate, the higher are the relative gains of workers in the socialization process. The joint evolution of culture and institutions displays a unique ergodic stationary state, $(\beta^* = 1, q^*)$: *extractive* institutions are undermined by their own inefficiency (due to the lack of commitment of the policy maker). The transition away from *extractive* institutions is inevitable, from any initial condition.[36] The complementarity between culture and institutions in this society is such

[36]The mechanism driving the dynamics of institutions is distinct from the one stressed by Acemoglu (2003),

Figure 4: Elites, workers and extractive institutions: Joint dynamics - phase diagram



that their joint dynamics, from any initial condition on the left of $q(\beta)$ (Region A in Figure 4), is monotonically increasing. Along these dynamical paths, the political power of workers β and their fraction in society q reinforce each other. Since the stationary state for β given q is constant ($= 1$), the cultural multiplier is 0 in this society (both the multiplier from any initial condition and the multiplier at the stationary state); see Proposition 8. Nonetheless, and most importantly, the multiplier of aggregate income in society is positive, from any initial condition on the left of $q(\beta)$. Indeed, aggregate income is the total income of workers (as the elites are a leisure class), $a^1(\beta, q)q$; per-capita income $a^1(\beta, q)$ is increasing in β and q and both the power β and fraction q of workers increase along these paths. The multiplier of aggregate income in society is then: $\frac{a^1(1,q^*)}{a^1(1,q_0)} \frac{q^*}{q_0}$.

Acemoglu and Robinson (2006, 2010) and Acemoglu, Johnson, and Robinson (2006). In our society, the transition is triggered independently of any technology on the part of the workers to threaten, e.g., by means of a revolution, the power of the aristocrats. Furthermore, in this society, extractive institutions are not stable, independently of the population distribution between workers and elites. In particular, this is the case even if the relative power of workers is unaffected by their relative size (or even relative income) in society.

4.1.1 The transition away from extractive institutions.

In the society we have just studied in the previous section, the elites' lack of commitment leads to very inefficient equilibrium outcomes when the elites themselves control the institutional set-up. As a consequence the joint dynamics of culture and institutions necessarily drive the society towards less extractive institutions where fiscal authority is devolved mostly (or even completely, in some parameters' configurations) to workers. In this section we study a simple extension of this society with a more articulate and interesting representation of the transition away from extractive institutions, one which delves more deeply into the cultural preferences and the incentives of the elites. In particular, while very stylized, this example will identify a fundamental role of the bourgeoisie in this process; foreshadowing either a representation of the formation of inclusive institutions protecting property rights in England (McCloskey, 2006, 2010, 2017) or else of the maintenance of slavery institutions in the Roman Empire (Schiavone, 1996).

Consider a society alike to the one studied in the previous section except that i) a fraction of the members of the elite hold a cultural trait which specifies work-ethic norms akin to those of the workers, rather than those of the leisure class;[37] ii) workers are not endowed with an initial endowment of resources, they can only consume off their production; iii) workers face a survival constraint, a minimum level of consumption necessary for survival. Furthermore, in the society we study in this section, taxes are not raised to finance the public good but are instead purely extractive, being redistributed pro-capita to the members of the elite.[38] As in the previous section, the elites, as a political group, have the power of taxing workers, but cannot commit ex-ante to a tax rate. In this society, however, their incentives and preferences are heterogeneous: the members of the elite who share work-ethic norms (the *bourgeois*) are more aligned with workers' interests than those who do not (the *aristocrats*). Depending on their distribution by cultural trait and the political control they exert on the fiscal authority in society, the elites might impose a tax rate such that workers are constrained to subsistence (an *extractive regime*).[39] The institutional dynamics of this economy will in general be *non-ergodic*, depending crucially on initial conditions. Only when the initial institutional set-up guarantees enough control on fiscal authority on the part of the workers, the institutional dynamics will tend to transition away from the *extractive regime*. Interestingly, this transition will generally induce the formation of a sizeable bourgeoisie. As well, it is also the case that a larger bourgeoisie at the initial conditions favors the transition away from the *extractive* regime.

The detailed analysis of this society follows. Workers, group $i = 1$, are in proportion $1 - \lambda$

[37] Note that in this society, therefore, political groups (workers and elites) are not aligned with cultural groups (bourgeois and aristocrats, inside the elite). The example is then a special case of the class of societies introduced in Section 5 and therefore follows the notational structure laid out there.

[38] This is not substantial to the analysis. It is just for the sake of variation.

[39] The survival constraint can be binding only for workers, as members of the elites are endowed with initial resources which we postulate are enough for survival.

and members of the elites, group $i = 2$, in proportion λ. Members of the elite carry one of two possible cultural traits, $j = a, b$: i) the *bourgeois*, in proportion $q^{2b} = q$ of the total elite size λ, have the same preferences as workers; ii) the *aristocrats* are instead in proportion $q^{2a} = 1 - q$ of the elite and have preferences with extreme disutility for work. All agents have preferences over a consumption good c^{ij} and labor effort a^{ij}, where $i = 1, 2$ indexes the group and j the cultural trait.[40] The production technology converts effort one-to-one in the consumption good. Let $\beta^1 = \beta$ denote the institutional weight of the workers, and p, the policy choice, represent the tax rate on workers' output, a^1. Let T denote the lump sum fiscal transfer received by each member of the elite, by budget balance. Let $\bar{c}$ denote the subsistence level required for survival, and assume that workers do not have initial resources (ie. $s^1 = 0$) while all members of the elite have an initial endowment $s^2 = s > \bar{c}$ The per-capita fiscal transfer to the members of the elite is set to balance the budget of the fiscal institutions: $T =: \frac{1-\lambda}{\lambda} p a^1$. Preferences are represented by the following utility functions, respectively for workers and elites:

$$u^1(a^1, T^1, p) = u\left(a^1(1-p)\right) + \theta^1 v(1 - a^1)$$

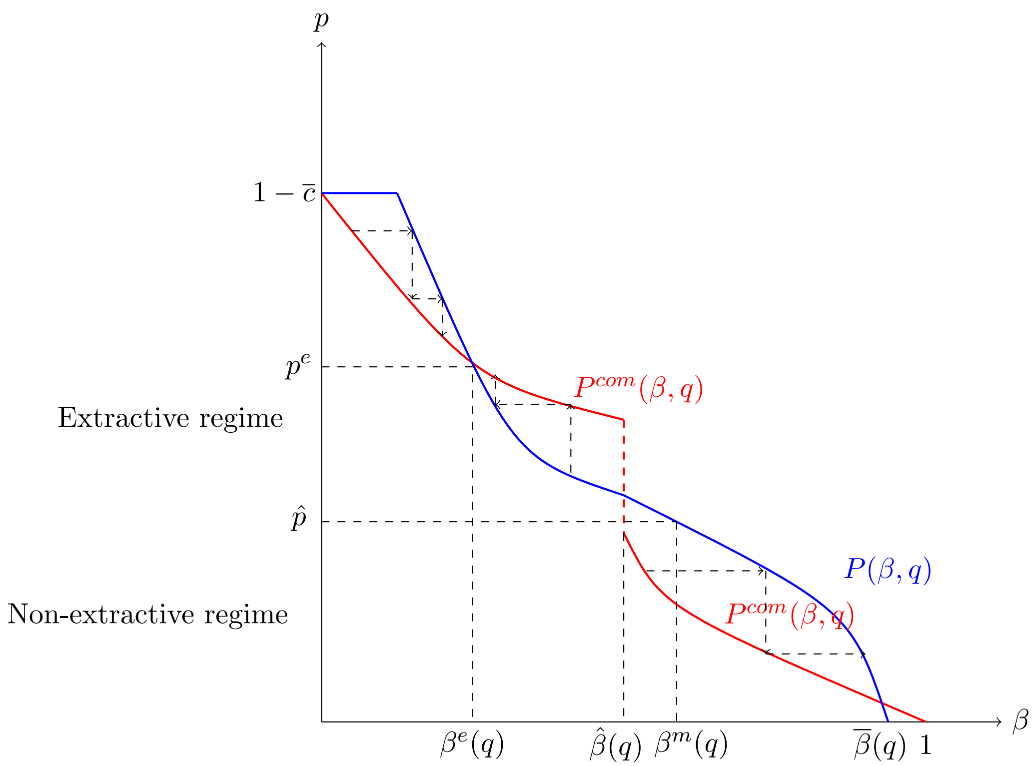
$$u^{2j}(a^{2j}, T^2, p) = u\left(a^{2j} + s + T\right) + \theta^{2j} v(1 - a^{2j})$$

Again we assume that the aristocrats have extreme preferences for leisure $\theta^{2a} > \frac{u'(s)}{v'(1)} > 1 = \theta^{2b} = \theta^1$ so that again they never work, $a^{2a} = 0$. A crucial aspect of this society consists in the fact that the labor effort exerted by workers, $a^1(p)$, is non-monotonic in the tax rate p, depending on whether the survival constraint is binding. When the tax p is smaller than a cutoff $\hat{p}$, the survival constraint is not binding and $a^1(p)$ is decreasing in p, because of the disincentive effects of the tax rate on effort. But when instead $p > \hat{p}$, the survival constraint is binding, $a^1(p) = \frac{\bar{c}}{1-p}$, and workers' labor effort increases with p to maintain survival. We call this, the *extractive* regime.

The *societal equilibrium policy* $p(\beta, q)$, and the *societal commitment policy* $p^{com}(\beta, q)$ are illustrated in Figure 5. When the institutional weight of the workers is low enough, the fiscal authorities, with or without commitment, would choose the *extractive regime* and tax the workers to a level that forces them to survival. Whether in this regime the commitment problem induces a tax which is too low or too high, that is, whether $P(\beta, q) = p^{com}(\beta, q) - p(\beta, q)$ is $\gtreqless 0$, depends on the balance of two effects. On one hand, as we noted, at survival higher taxes increase the effort of workers. A fiscal authority lacking commitment would not internalize this effect, inducing a *societal equilibrium policy* $p(\beta, q)$ lower than the *societal equilibrium policy with commitment* policy $p^{com}(\beta, q)$. On the other hand, in this regime it is also the case that taxes p cause a distortion on workers' welfare which tends to make $p(\beta, q)$ too high compared to $p^{com}(\beta, q)$ and is not internalized at the social equilibrium. When β is very low, this distortionary effect on the workers' welfare dominates. But as β increases the first effect tends to dominate and $p^{com}(\beta, q)$ and $p(\beta, q)$ cross as in the figure. In the *non-extractive regime*, instead, when workers are not

[40] Abusing notation the apex j is omitted for workers, $i = 1$, since they are culturally homogeneous.

Figure 5: Transition away from extractive institutions: Equilibrium policies and institutional dynamics



at survival as in the society studied in the previous section, a fiscal authority with commitment would internalize the negative effect of taxes on the tax base and $p^{com}(\beta, q) < p(\beta, q)$.

The institutional dynamics, fixing a cultural distribution $0 < q < 1$, depends on the initial condition. For all initial value $\beta_0 < \hat{\beta}(q)$, the dynamics converge to a unique steady state $\beta = \beta^e(q)$ and the society ends up in an *extractive* regime with low political representation of the workers who are maintained at their survival constraint by extractive taxation on the part of the elites.[41] Conversely for initial values $\beta_0 > \hat{\beta}(q)$, the institutional dynamics are very different. The weight of the workers on the institutional setting converge to $\beta = \beta^*$,[42] with no taxation, in a non-extractive regime.[43] Interestingly, for initial conditions β_0 between $\hat{\beta}(q)$ and $\overline{\beta}(q)$ the society will move away from the extractive into the non-extractive regime.

Given the long-run dynastic dynamics we are considering in this example, we adopt the inter-generational cultural transmission mechanism of Bisin and Verdier (2001). The dynamics of cultural evolution within the elite are then driven by the relative incentives to socialization $\Delta V^b(p)/\Delta V^a(p)$, which iare generally decreasing in p. Indeed, aristocratic norms are more likely to be transmitted than those of the bourgeoisie the larger the rents of the elites. Since equilibrium taxation is a decreasing function of the institutional weight of workers, β, the more fiscal authority the workers possess in society, the larger the diffusion of (the norms of) the bourgeoisie inside the elite, and hence in society.

The joint dynamics of culture and institutions in this society will in general be *non-ergodic*: which stationary state they will converge to in the long-run depends on initial conditions; see the phase diagram in Figure 6. There are two types of stationary states of the dynamical system: the stationary *extractive* state (β^e, q^e), and steady states where workers are not taxed with $\beta \geq \beta^*$.[44] A transition away from *extractive* institutions is not inevitable in this society, as higher taxes do not decrease the fiscal rents of the elites when workers are at or around the survival constraint. *Extractive* institutions are therefore not any more undermined by their own inefficiency and could be supported in the long-run. Whether *extractive* institutions are supported in the long-run, or whether the dynamics transition away, depends on the political control the elites exert on the fiscal authority in society, but also on their distribution by cultural trait (that is, the relative size of the bourgeoisie, which is partly aligned with workers' interests). The basin of attraction of the extractive state (β^e, q^e) is the whole region below $\hat{\beta}(q)$. It is of interest to note that culture and institutions are substitutes at the extractive state (β^e, q^e) and hence oscillatory dynamics cannot

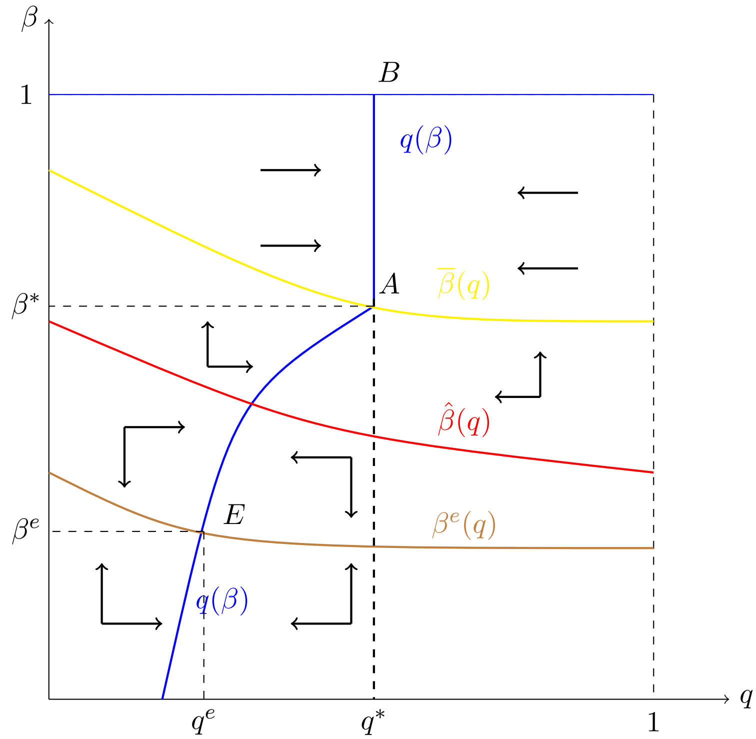
[41] Interestingly in the *extractive* regime, higher taxation may actually increase the efficiency of the rent extraction process as the survival constraint prevents the traditional disincentives on labor supply to kick in. This local effect is arguably instrumental in maintaining such an extractive regime for workers. This is reminiscent of an argument in Clark (2007, chapter 2), suggesting that policies that would otherwise appear as having inefficiency costs in a non extractive world, on the contrary may find some efficiency rationale under extractive conditions.

[42] As well it stays constant at any value $\beta \geq \beta^*$.

[43] The dynamics from $\beta_0 = \hat{\beta}(q)$ are indetermined.

[44] These steady states are given by all values $\beta \geq \beta^*$ and $q = q^*$.

Figure 6: Transition away from extractive institutions: Phase diagram



be ruled out, from any initial conditions in this region.

When the initial institutional set-up ensures enough control on fiscal authority on the part of the workers, the dynamics will tend to transition away from the *extractive regime*. But a larger bourgeoisie at the initial conditions also favors the transition away from the *extractive* regime. Formally, the basin of attraction of stationary states in which the workers are not taxed, i.e., characterized by $\beta \geq \beta^*$ and $q = q^*$), comprises the whole region (strictly) above the line $\widehat{\beta}(q)$. It is larger in β for higher q and it is also larger in q for higher β. For any initial conditions in this region, the dynamics of culture and institutions display a transition away from an extractive society, where workers are taxed and kept at survival, into an inclusive society where workers consume over and above survival. Furthermore, once reached an inclusive society, the elites delegate political power to the workers, and taxes decrease. Along this dynamic path, which converges to the stationary states where workers are not taxed, the size of the bourgeoisie grows monotonically.

In this society, as in the one introduced in the previous section, the institutional stationary states with no taxation are constant in q, and the cultural multiplier is 0 (both the multiplier from any initial condition and the multiplier at the stationary state); see Proposition 8. However, for $q_0 < q^*$, given that the bourgeoisie works and hence, opposite to the aristocratic leisure class is productive, the multiplier of aggregate income is positive.

4.2 Civic Capital and institutions

In this example we study *civic capital*, that is, a society where a fraction of the population is endowed with intrinsic motivations which induce actions that may have beneficial effects on other members of the society.[45] We study in particular conditions under which the cultural and institutional dynamics in this society favors or hinders the spread of *civic capital*.[46]

Consider a society populated by two groups, workers and members of the elite, as in the previous example. Both workers and members of the elite are endowed with the same technology which transforms labor into private consumption goods. Fiscal institutions collect lump-sum taxes to finance the provision of a public good, whose consumption is valued by both groups. The provision of the public good, however, creates opportunities for corruption that benefit exclusively the elite. We can therefore think of the elite as of a *caste of bureaucrats*. The preferences of a fraction of the workers are shaped by a *civic culture*, which motivates them to exert a participation effort, complementary to the provision of the public good, as well as a monitoring effort to fight corruption, costly to the elite. *Civic participation* involves e.g., contributing privately to public goods, creating social associations, volunteering in social activities. *Civic control* creates transparency, by monitoring the government in its public good provision process.

In this society, therefore, public good provision is associated to different externalities on society. On the one hand, it stimulates the civic participation of a fraction of the workers, a positive externality on society as a whole. On the other hand, public good provision induces corruption and the reaction of a fraction of workers against it, a positive externality on workers with no civic capital and a negative externality on the elite. Institutions lack commitment; that is, fiscal authorities choose lump-sum taxes to finance public good provision without internalizing the effects of civic capital in society. Public good provision can be larger or smaller at equilibrium than the efficient level, depending on whether the positive or the negative externalities of civic capital dominate. As a consequence, the institutional dynamics lead to a stationary balanced allocation of power between workers and the elite.

Most interestingly, in this society, culture and institutions may act as substitutes. Indeed, on the one hand, the incentives to transmit civic culture are generally increasing in the political representation of workers in society; on the other hand, the larger is the spread of civic capital in the population of workers, the smaller are the incentives to design institutional changes devolving power to the workers, as the beneficial effects of civic capital are already present. In this society,

[45] We use *civic culture* as the individual trait and *civic capital* as the societal characteristic - though this distinction may appear hard to distinguish in the text.

[46] See also Ticchi,Verdier and Vindigni (2013) and Besley and Persson (2019) for specific analyses of the interactions between political culture and political institutions.

therefore, an exogenous institutional change which endows with more political power citizens could see its effects mitigated by the induced cultural dynamics associated to the spread of civic capital in the population.

Formally, let workers be group $i = 1$ and the elite be $i = 2$, in fractions $\lambda^1 = 1 - \lambda^2 = \lambda$. Workers belong to one of two cultural groups. The first, $j = c$ in proportion q, is characterized by preferences endowed by civic culture; the second group is passive with respect to the civic society, $j = p$. All individuals (workers and elite) are endowed with a fixed amount of resources, s. Lump-sum taxes are raised to finance public expenditures, g. In the process of providing for a public good, a fraction $\mu > 0$ of public expenditures leaks into corruption, generating diverted rents $T = \mu g$ that benefit exclusively the members of the elite. The residual share of public good expenditures is used to provide public good, $G = (1 - \mu)\, g$. Workers can exert two types of efforts, *civic participation* and *civic control*. Let worker's j participation (resp. control) effort be denoted e^{1j} (resp. a^{1j}). Societal civic participation effort is then $E = \lambda \left[q \cdot e^{1c} + (1 - q) \cdot e^{1p}\right]$, while societal civic control effort is $A = \lambda \left[q \cdot a^{1c} + (1 - q) \cdot a^{1p}\right]$. Societal civic participation effort E produces a society wide externality which augments each individual's endowment by $\kappa \cdot E$, $\kappa > 0$. Societal civic control effort A increases the transaction costs associated to corruption activities: the consumption associated to T units of diverted rents is $(1 - \theta A)\, T$, with $0 < \theta < 1$.[47] The government policy is total public expenditures g, financed by lump-sum taxes in the same amount. Workers' preferences are as follows:

$$\begin{aligned} U^{1c}\left(c^{1c}, G, a^{1c}, e^{1c}, T\right) &= c^{1c} + v(G) - (\alpha \cdot T)\,(1 - a^{1c}) - C(a^{1c}) + G \cdot e^{1c} - \Phi(e^{1c}) \\ U^{1p}\left(c^{1p}, G, a^{1p}, e^{1p}, T\right) &= c^{1p} + v(G) - C(a^{1p}) - \Phi\left(e^{1p}\right); \end{aligned}$$

where $c^{1c} + v(G)$ is the direct utility of private consumption and the public good; $-(\alpha \cdot T)\,(1 - a^{1c})$ is the intrinsic motivation for civic control; $C(a^{1c})$ is the utility cost of undertaking civic control; $G \cdot e^{1c}$ is the intrinsic motivation to contribute e^{1c} to civic participation; while $\Phi(e^{1c})$ is the disutility cost of civic participation.[48] Members of the elite have standard preferences over consumption and the public good:

$$U^2(c^2, G) = c^2 + v(G).$$

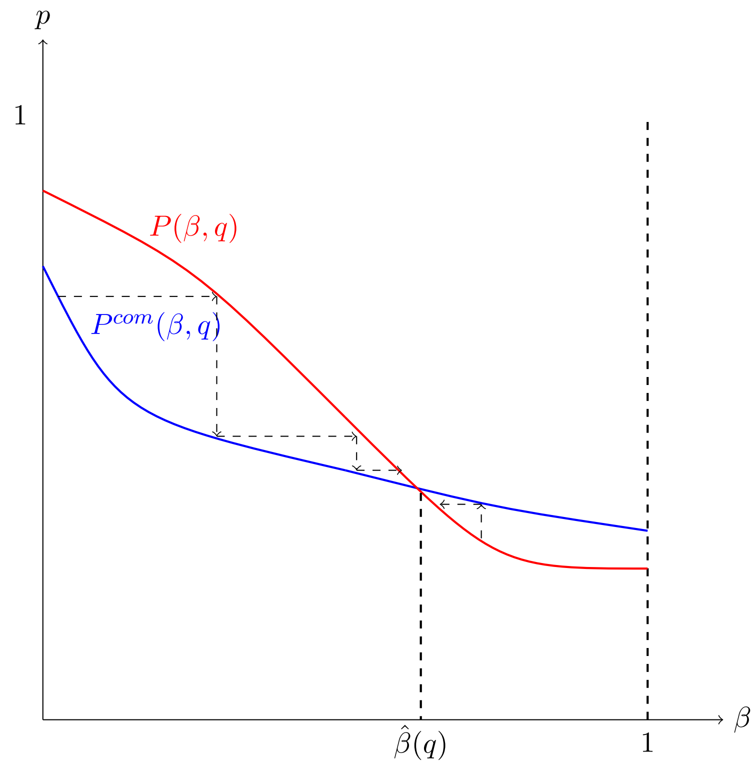
Policy choice $p = g$ depends on the workers' efforts only through $\theta \cdot A$ and $\kappa \cdot E$. Therefore, since the contribution of each worker effort to societal efforts E and A is negligible, passive workers always choose not to exert any effort, $a^{1p} = e^{1p} = 0$, and workers with civic culture contribute according to their intrinsic motivations. In fact, since both G and T increase in g, e^{1c} and a^{1c} also increase in g.

Under some reasonable regularity conditions, $p\,(\beta, q)$ and $p^{com}\,(\beta, q)$ are as in Figure 8: downward sloping in β, for any q. More specifically, when the character of civic culture is not too

[47] Effort costs are normalized so that $\theta A < 1$.

[48] See the Online Appendix for details, assumptions, and functional forms.

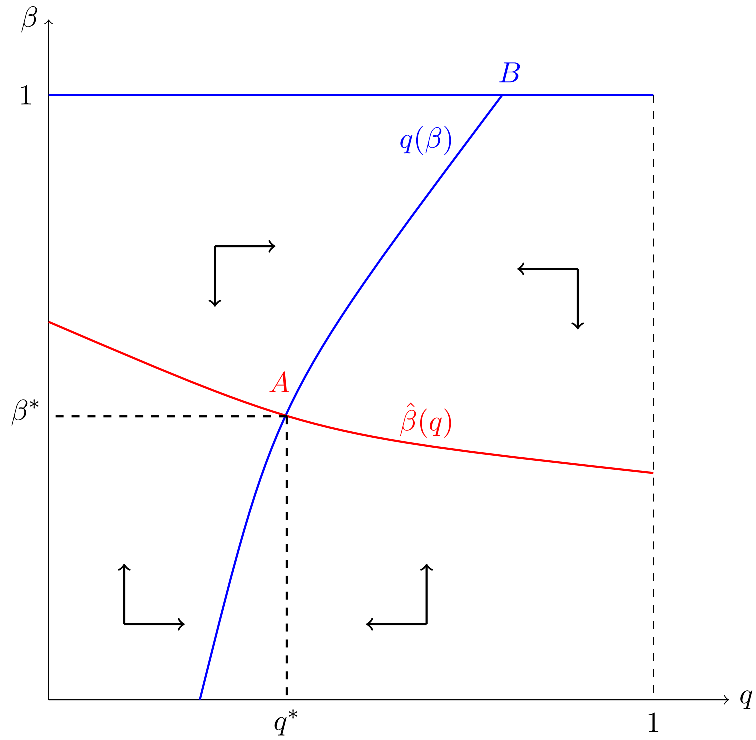
Figure 7: Civic capital and institutions: Institutional dynamics



unbalanced in terms of civic participation, that is, workers with civic culture are less in favor of large public expenditures than the elite. As a consequence, an increase in β would tend to reduce the size of the public expenditures at the *societal equilibrium*, $p(\beta, q)$, and the *societal commitment*, $p^{com}(\beta, q)$. For the same reason, at a given value of β, an increase in the fraction of workers with civic culture, q, would have the same effect on public expenditures. Most importantly, $p(\beta, q)$ crosses $p^{com}(\beta, q)$ from above at some interior point $\widehat{\beta}(q)$. Indeed, $p^{com}(\beta, q)$ is the policy choice once all externalities in society are internalized. But the negative externality, via $\theta \cdot A$, is born out only by elite members, while the positive externality, via E, is enjoyed by the whole society. As a consequence, when the political power of the elite is large (i.e., β small), internalizing the negative externality dominates the society's political objectives and $p^{com}(\beta, q) < p(\beta, q)$. Conversely, when the weight of the elite is small, internalizing the positive externality dominates and consequently $p^{com}(\beta, q) > p(\beta, q)$. For all initial values β_0. the institutional dynamics converge to a unique steady state $\beta = \widehat{\beta}(q)$ and political power is shared between the workers and the elite; see Figure 7.

Importantly, in this society, the political power of workers at the stationary state, $\widehat{\beta}(q)$, is decreasing in the predominance of civic capital, q. This is because societal civic control A increases in q and A substitutes for formal political power. More in detail, at the stationary state $\widehat{\beta}(q)$, $p^{com}\left(\widehat{\beta},q\right)=p\left(\widehat{\beta},q\right)$ and the positive and negative externalities associated to public expenditures balance out at the margin in the government objective function. An increase in q would lead to fewer public expenditures, as workers with civic culture are more concerned than the rest of society by corruption. To restore the equilibrium, institutional dynamics move then in the direction of re-introducing larger public expenditures and hence of reducing the political power of workers, β.

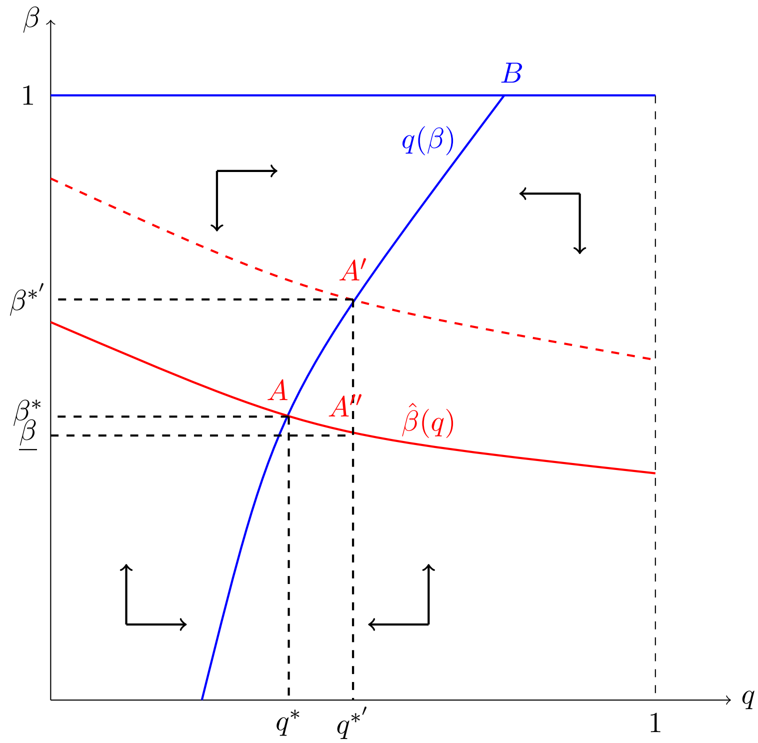
Figure 8: Civic capital and institutions: Phase diagram



The elite is culturally homogenous and hence displays no cultural dynamics. The cultural dynamics within workers are determined by the relative incentives to transmit civic culture, $\Delta V^{1c}(p)/\Delta V^{1p}(p)$, as they depend on the equilibrium policy instrument p. When civic participation e^{1c} is less sensitive to public good provision than civic monitoring a^{1c}, $\Delta V^{1c}(p)/\Delta V^{1p}(p)$ is decreasing in p. As the societal equilibrium $p\left(\beta,q\right)$ is itself a decreasing function of β and q, the relative incentives to transmit civic culture increase with both β and q in society. As a consequence, $q\left(\beta\right)$ is upward sloping in β: the formal delegation of power to the workers tends to induce a larger diffusion of civic capital between workers; see Figure 7.

The joint evolution of culture and institutions is also illustrated in Figure 8. The stationary state of the joint dynamics is (β^{*},q^{*}). At (β^{*},q^{*}), the two manifolds $\beta(q)$ and $q(\beta)$ have slopes of opposite signs. As a consequence, culture and institutions are substitute in this society and the cultural multiplier is negative (see Proposition 8): the effect of an exogenous shock which changes the political power of workers in some direction would be mitigated by the ensuing cultural dynamics. Figure 9 describes the effects of an increase in the coefficient κ, which, other things equal, increases the positive externality associated to civic participation E. A change in κ triggers a higher demand for public expenditures, and therefore some institutional dynamics biased against the workers' group. This institutional change in turn reduces the relative incentives to transmit civic culture and leads to a reduction of q. As civic capital is reduced, there is less civic control effort against corruption in society. This in turn calls for some institutional change returning some formal power to workers, mitigating therefore the initial institutional impact of the shock to κ.

Figure 9: Civic capital and institutions: Comparative dynamics



Interestingly, depending on the relative speeds of the dynamics of culture and institutions, the dynamics of adjustment to the shock may not be monotonic. Suppose for instance that institutions adjust much faster than culture, so that the adjustment dynamics lies on $\widehat{\beta}(q)$. In this case, the shock on κ, after having induced β to jump downwards (with q constant at q^*), has q decrease and β increase along the adjustment path.

5 Conclusions and implications for empirical studies

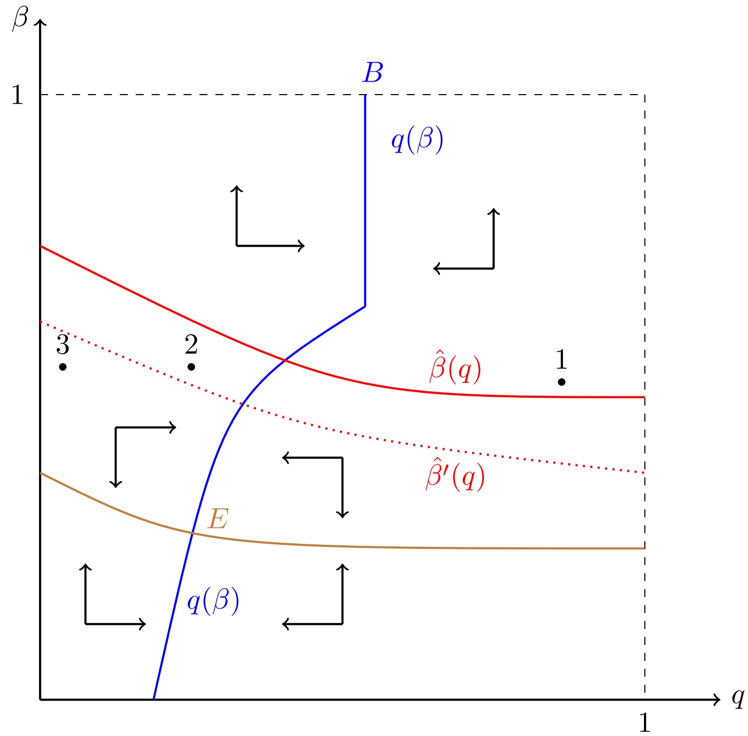
In this paper we develop theoretical and empirical tools for the analysis of the effects of culture and institutions on economic variables of interest, notably, long-run economic activity. In our view, the theoretical model we develop, the concepts of cultural and institutional multipliers we introduce, and the examples we construct to study the political economy of elites and civic society, provide some needed structure complementing the role of causal analysis in *persistence studies* with regards to interesting historical phenomena.

Depending on whether culture and institutions are dynamic complements or substitutes, exogenous historical shocks propagating over the joint dynamics induced by institutions and culture may have magnified or mitigated effects on long-run socio-economic outcomes. This type of analysis identifies the extent of the comparative dynamic bias that is generated by neglecting one of the two dynamics, when the other one is affected by an exogenous shock. We surmise that this is of first order importance when studying the long-run effects of historical shocks. Along these lines, while no empirical study has yet attempted to estimate the size of the cultural and the institutional multipliers in the context of *persistence studies*, several papers provide explicit quantitative evidence about their sign; that is, they document whether culture and institutions acted as complements or substitutes. Lowes, Nunn, Robinson, and Weigal (2017) find evidence for substitution along the development of the Kuba Kingdom in Central Africa in the 17th century. Lowes and Monteiro (2017) also find substitution in their study of the rubber extraction system in the Congo Free State during the colonial era. On the other hand, Dell (2010)'s study of forced mining labor in Peru and Bolivia in the 16th century provides suggestive evidence of complementarity.

Our approach also highlights that in general the joint evolution of culture and institutions has some highly non-linear components. Non-linearities are at the root of the interesting dynamics we illustrate in the specific example societies in Section 4, from sensitivity of equilibrium trajectories to initial conditions to thresholds effects and non-monotonicity of cultural and institutional changes over transition paths. These phenomena appear indeed quite consistent with the great diversity of development trajectories encountered across the world and in time. Most importantly, they open a relevant role for structural models in empirical studies of the determinants of historical phenomena, beyond the standard causal identification analysis. As an illustration, consider the following conceptual experiment, relying on our analysis of the transition away from extractive institutions in Section 4.1.1. Recall the phase diagram of the dynamics in this society, in Figure 6, which we report again in Figure 10. E and B are two stationary states,[49] with distinct basins of attraction whose boundary is delineated by the red line $\hat{\beta}(q)$. Consider three countries, with different initial conditions in terms of culture, denoted $1, 2, 3$ in the figure. Consider an ex-

[49] More precisely, the whole vertical line from B is a stationary state, with the same properties as B. We disregard this for simplicity in the discussion below.

Figure 10: (Local) Average Treatment Effects of institutional change



Country 1: Always taker; Country 2: Complier; Country 3: Never taker

ogenous shock to institutional set-up of the socio-economic environment affecting the dynamical system, shifting the red line down to the red-dotted line. This shock has no qualitative effect on the dynamics of country 1 and 3, the first converging to the stationary state E and the other to B. The shock, on the other hand, leads Country 2 to converge to E rather than B. In the terminology of causal analysis, countries like 1 and 3 are called, respectively, *always takers* and *never takers*; countries like 2 are instead *compliers*. An empirical cross-sectional exercise estimating the effects of this shock on a long-run variable linked to culture and institutions would identify the *Average Treatment Effect* limited to compliers; that is the *Local Average Treatment Effect*, which will generally differ from the Average Treatment Effect when countries are heterogeneous. In this example, the structural model (represented by the phase diagram) suggests the existence of an initial threshold determining the long run dynamic path of societies which, when disregarded, induces the econometrician to a misleading interpretation of the effects of an institutional shock to the socio-economic environment.[50]

[50] See Bisin and Moro (2021) for a discussion of these aspects in the context of various *persistence studies* in the literature; and Casey and Klemp (2021) for a bias correction method to instrumental variable estimators in related contexts.

References

Abad, L.Arroyo and N. Maurer (2021), “History Never Really Says Goodbye: A Critical Review ofthe Persistence Literature,” *Journal of Historical Political Economy*, 1(1), 31–68.

Acemoglu, D. (2003): ”Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics,” *Journal of Comparative Economics*, 31, 620-52.

Acemoglu, D., G. Egorov, and K. Sonin (2015): ”Political Economy in a Changing World,” *Journal of Political Economy*, 123(5), pp. 1038-1086.

Acemoglu, D., G. Egorov, and K. Sonin (2021): ”Institutional Change and Institutional Persistence”, in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: Elsevier North Holland, 365-389.

Acemoglu D. and M. Jackson (2015), ”History, Expectations, and Leadership in the Evolution of Social Norms”, *Review of Economic Studies*, 82, 2, 423-456

Acemoglu, D., S. Johnson and J.A. Robinson (2001): ”The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, 91, 1369-1401.

Acemoglu, D., S. Johnson and J.A. Robinson (2006): ”Institutions as the Fondamental Cause of Long-Run Growth,” in P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, Amsterdam, Elsevier.

Acemoglu D., and J.A. Robinson, (2000): ”Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective”, *The Quarterly Journal of Economics*, vol. 115, 4, pp. 1167-1199.

Acemoglu D., and J.A. Robinson, (2001), ”A Theory of Political Transitions”, *The American Economic Review,* Vol. 91, No. 4, pp. 938-963.

Acemoglu, D. and J.A. Robinson (2006): *Economic Origins of Dictatorship and Democracy*, Cambridge, Cambridge University Press.

Acemoglu, D. and J.A. Robinson (2010): *Why Nations Fail*, New York, Crown Publishers.

Acemoglu, D. and J.A. Robinson (2021); ”Culture, Institutions and Social Equilibria: A Framework”, NBER WP n°28832 .

Alesina, A. and P. Giuliano (2015): ”Culture and Institutions,” *Journal of Economic Literature*, 53(4), 898-944.

Alger, I., and J.W. Weibull (2013): Homo moralis Preference evolution under incomplete information and assortative matching, *Econometrica* 8, 2269-2302.

Almagro, M. and D. Andres-Cerezo (2020): "The Construction of National Identities," *Theoretical Economics*, 15 (2), 763-810.

Aghion P., I Algan, P.Cahuc and A. Schleifer, (2010), "Regulation and Distrust", *The Quarterly Journal of Economics*, 125, 3, 1015–1049.

Aron, R. (1950a): "Social Structure and the Ruling Class: Part 1," *The British Journal of Sociology*, 1(1), 1-16.

Aron, R. (1950b): "Social Structure and the Ruling Class: Part 2," *The British Journal of Sociology*, 1(2), 126-143.

Arroyo-Abad L. and J. Luiten van Zanden (2016), "Growth under Extractive Institutions? Latin American Per Capita GDP in Colonial Times, *The Journal of Economic History,* 76,4, 1182-1215.

Belloc M. and S. Bowles (2013): "The Persistence of Inferior Cultural-Institutional Conventions", *American Economic Review: Papers and Proceedings* 103(3), 1-7.

Belloc M. and S. Bowles (2017): "Persistence and Change in Culture and Institutions under Autarchy, Trade, and Factor Mobility", *American Economic Journal: Microeconomics* 9(4), 245-276.

Besley, T. (2017), "Aspirations and the Political Economy of Inequality," O*xford Economic Papers,* 69(1), 1-35.

Besley, T. (2020), "State Capacity, Reciprocity, and the Social Contract," *Econometrica*, 88(4), 1307-1335.

Besley, T. and M. Ghatak (2017), "The Evolution of Motivation", LSE Mimeo.

Besley, T., and T. Persson (2019), "Democratic Values and Institutions", *American Economic Review Insights*, 1(1), 59-76.

Besley, T., and T. Persson (2020), "Organizational Dynamics: Culture, Design, and Performance", LSE mimeo.

Benabou R., D. Ticchi and A. Vindigni (2015), "Religion and Innovation", *American Economic Review*, 105, 5, 346-351.

Bidner C. and P. Francois, (2011), "Cultivating Trust: Norms, Institutions and the Implications of Scale," *Economic Journal,* 121(555), pp. 1097-1129, 09.

Bisin, A. and A. Moro (2021)': "LATE for History", in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: ElsevierNorth Holland.

Bisin A, J. Rubin, A. Seror and T.Verdier (2021), "Culture, Institutions and The Long Divergence", NBER WP n°28488 and CEPR DP n°15802.

Bisin, A and T. Verdier (1998): "On the Cultural Transmission of Preferences for Social Status," *Journal of Public Economics,* 70, 75–97.

Bisin, A and T. Verdier (2000a): "Beyond the Melting Pot: Cultural Transmission, Marriage and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics,* 115, 955-88.

Bisin, A. and T. Verdier (2000b): "Models of Cultural Transmission, Voting and Political Ideology," *European Journal of Political Economy*, 16, 5-29.

Bisin, A and T. Verdier (2001): "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97, 298–319.

Bisin, A. and T. Verdier (2010): "The Economics of Cultural Transmission," in Benhabib, J., A. Bisin, M. Jackson (eds.), *Handbook of Social Economics*, Amsterdam, Elsevier.

Bisin, A. and T. Verdier (2021): "Phase Diagrams in Historical Economics: Culture and Institutions, in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: ElsevierNorth Holland.

Boranbay, S. and C. Guerriero (2019): "Endogenous (In)Formal Institutions," *Journal of Comparative Economics*, 47, 4, 921-945.

Bogart D. and L. Chaudhary (2019), "Extractive institutions? Investor returns to Indian railway companies in the age of high imperialism ", Journal of Institutional Economics, 15, 5, 751 - 774.

Boyd, Robert, and Peter Richerson (1985): *Culture and the Evolutionary Process,* Chicago, IL, University of Chicago Press.

Bjornerstedt J, and J. Weibull (1996): Nash Equilibrium and Evolution by Imitation, in *The Rational Foundations of Economic Behaviour*, ed. K.J. Arrow, E. Colombatto, M. Perlman and C. Schmidt, 155- 181, Mew York: St. Martin's Press.

Bowles S., J-K Choi, S-H Hwang and S. Naidu, (2021), " How Institutions and Cultures Change: an Evolutionary Perspective", in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: Elsevier North Holland, 391-433.

Cantoni, D. and N. Yuchtman (2021): "Historical Natural Experiments: Bridging Economics and Economic History," in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: ElsevierNorth Holland.

Carvalho, J.P.,(2016), "Identity-Based Organizations", *American Economic Review,* 106(5), 410–414.

Carvalho, J.P., and C. Dippel (2016): "Elite Identity and Political Accountability: A Tale of Ten Islands," NBER Working Paper 22777.

Carvalho, J.P. and M. Sacks (2021), "The Economics of Religious Communities", forthcoming *Journal of Public Economics.*

Casey, G. and M. Klemp (2021): "Instrumental Variables in the Long Run," *Journal of Development Economics*, forthcoming.

Cavalli Sforza, L.L. and M. Feldman (1973): "Cultural Versus Biological Inheritance: Phenotypic Transmission from Parent to Children," *American Journal of Human Genetics*, 25, 618-37.

Cavalli Sforza L.L. and M. Feldman (1981): *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, NJ, Princeton University Press.

Chen, Y., H. Wang, S. Yan (2014): "The Long-Term Effects of Protestant Activities in China," mimeo, Peking Un.

Clark G. (2007), *A Farewell to Alms: A Brief Economic History of the World (The Princeton Economic History of the Western World.* Princeton, NJ, Princeton University Press.

Conley J. and A. Temimi (2001), "Endogenous Enfranchisement When Groups' Preferences Conflict", *Journal of Political Economy*, Vol. 109, (1), pp. 79-102.

Davis, L.S., and C.R. Williamson (2016), "Culture and the regulation of entry," *Journal of Comparative Economics*, 44(4), 1055-1083.

Dell, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78(6), pp. 1863-1903.

Dell M., and B.A. Olken (2020), "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java", *The Review of Economic Studies,* 87, 1, 164–203.

Diamond J. (2005), *Collapse: How Societies Choose to Fail or Succeed*, Penguin Books, New York.

Doepke M. and F. Zilibotti (2008): "Occupational Choice and the Spirit of Capitalism," *Quarterly Journal of Economics*, 123(2), 747-793.

Engerman, S. and K. Sokoloff (2002), 'Factor Endowments, Inequality, and Paths of Development among New World Economies', NBER Working Paper No. 9259

Gorodnichenko, Y. and G. Roland (2011): "Which Dimensions of Culture Matter for Long-Run Growth?," *American Economic Review, Papers & Proceedings*, 101(3), 492-98.

Gorodnichenko, Y. and G. Roland (2017): "Culture, Institutions, and the Wealth of Nations" , Review of Economics and Statistics, 99, 3, 402-416.

Gorodnichenko, Y. and G. Roland (2021), "Culture, Institutions and Democratization", *Public Choice*, 187(2), 165–195.

Gradstein M. (2007), "Inequality, Democracy and the Protection of Property Rights", *The Economic Journal,* 117, 516, 252-269.

Gradstein M. (2008), "Institutional Traps and Economic Growth", *International Economic Review*, 49,3, 1043-1066

Gramsci, A. (1951): *La Questione Meridionale,* edited by F. De Felice and V. Parlato, Rome, Editori Riuniti.

Greif, A. and G. Tabellini (2010): "Cultural and Institutional Bifurcation: China and Europe Compared," *American Economic Review, Papers & Proceedings*, 100:2, 1-10.

Greif, A. and G. Tabellini (2017): The clan and the corporation: Sustaining cooperation in China and Europe, *Journal of Comparative Economics*, 45, 1, 1-35

Guimaraes, B. and K.D. Sheedy (2016): "Guarding the Guardians", *Economic Journal*, forthcoming.

Guiso, L. P. Sapienza, and L. Zingales (2016): "Long-Term Persistence," Journal of the European Economic Association, vol 14, 6, pp. 1401-1436.

Guiso, L. P. Sapienza, and L. Zingales (2008): "Social Capital and Good Culture," Marshall Lecture, *Journal of the European Economic Association*, 6(2-3), pp. 295-320.

Güth W., (1995): An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives, I*International Journal of Game Theory*, 24; 323–344.

Güth W., and M. Yaari (1992): An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game. In: U. Witt (ed), *Explaining process and change—Approaches to evolutionary economics.* University of Michigan Press, Ann Arbor.

Hatcher, A. (2002): *Algebraic Topology*, Cambridge, Cambridge University Press.

Hauk, E and H. Mueller (2013), 'Cultural Leaders and the Clash of Civilizations', Journal of Conflict Resolution, 59,3, 367-400.

Helbing, D. (1992). "A Mathematical Model for Behavioral Changes by Pair Interactions", in Haag, G., Mueller, U.,and Troitzsch, K. G., (eds), *Economic Evolution and Demographic Change: Formal Models in Social Sciences*, 330–348. Springer, Berlin.

Huillery, E. (2009), 'History Matters: The Long-Term impact of Colonial Public Investments in French West Africa', *American Economic Journal - Applied Economics,*1(2), 176-215.

Hofbauer, J. (1995),." Imitation Dynamics for Games", Unpublished manuscript, University of Vienna.

Iyer L. (2010), "Direct versus Indirect Colonial Rule in India: Long-Term Consequences", *The Review of Economics and Statistics*, 92, 4, 693-713.

Iygun, M., J. Rubin, and A. Seror (2021): "A Theory of Cultural Revivals," *European Economic Review*, 135, 103734

Lagunoff, R. (2009): "Dynamic stability and reform of political institutions," *Games and Economic Behavior*, vol. 67(2), pages 569-583.

Landes, D.S. (1985):*The Wealth and Poverty of Nations: Why Some Are so Rich and Some so Poor*, New York, W.W. Norton and Co.

Lange M., J. Mahoney and M. vom Hau (2006), "Colonialism and Development: A Comparative Analysis of Spanish and British Colonies", *American Journal of Sociology*, 111, . 5, 1412-1462

Levine D. and S. Modica, (2021),"State Power and Conflict Driven Evolution", in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: Elsevier North Holland, 435-462.

Lindbeck A. (1995): "Hazardous Welfare State Dynamics," *American Economic Review*, 4, 9-15.

Lizzeri, A. and N. Persico (2004): "Why Did the Elite Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's Age of Reform"," *Quarterly Journal of Economics*, 707-65.

Lowes, S. and E. Monteiro (2017): "Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State," mimeo, Harvard University.

Lowes, S., N. Nunn, J.A. Robinson, and J. Weigel (2017): "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85, 1065-1091..

Mahoney, J., and K. Thelen (Eds) (2010): *Explaining Institutional Change –Ambiguity, Agency, and Power*, Cambridge University Press.

McCloskey, D.N. (2006): *The Bourgeois Virtues: Ethics for an Age of Commerce,* Chicago, University of Chicago Press.

McCloskey, D.N. (2010): *Bourgeois Dignity: Why Economics Can't Explain the Modern World,* Chicago, University of Chicago Press.

McCloskey, D.N. (2017): *Bourgeois DEquality: How Ideas, Not Capital or Institutions, Enriched the World,* Chicago, University of Chicago Press.

Michalopoulos, S. and E. Papaioannou (2013), 'Pre-Colonial Ethnic institutions and Contemporary African Development', *Econometrica* 81(1): 113-152.

Michalopoulos, S. and E. Papaioannou (2016), 'The Long-Run Effects of the Scramble for Africa', *American Economic Review* 106(7): 1802-1848.

Milnor, J.W. (1965)" *Topology from a Differentiable Viewpoint*, Charlottesville, University Press of Virginia [reprinted in Princeton Landmarks in Mathematics and Physics, Priceton, Princeton University Press, 1997].

Mokyr J. (2016), *A Culture of Growth: The Origins of the Modern Economy,* Princeton, Princeton University Press

Nunn, N. (2007), 'Historical Legacies: A Model Linking Africa's Past to its Current Underdevelopment', *Journal of Development Economics,* 83: 157-175.

Nunn, N. (2008), 'The Long Term Effects of Africa's Slave Trades', *Quarterly Journal of Economics,* 123: 139-176.

Nunn N. and L. Wantchekon, (2011), "The Slave Trade and the Origins of Mistrust in Africa", *American Economic Review,*101 (7): 3221-3252

Ogilvie S., (2021), "Thinking carefully about inclusiveness: evidence from European guilds", *Journal of Institutional Economics*, 17, 2, 185 - 200.

Ogilvie S. and A.W. Carus, (2014), " Institutions and Economic Growth in Historical Perspective", in Aghion P. and S. Durlauf (eds), *Handbook of Economic Growth,* vol. 2, 403-513.

Paredes Fuentes S. (2016), "Origins of Institutions and Inequality in Latin America", University of Warwick Mimeo.

Pareto, V. (1901): *Un'applicazione di Teorie Sociologiche*, *Rivista Italiana di Sociologia*, 402-456. [English translation (1968): *The Rise and Fall of Elites: An Application of Theoretical Sociology*, London: The Bedminster Press.

Persson T. and G. Tabellini (2021), "Culture, Institutions and Policy", in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: ElsevierNorth Holland,.463-489.

Prummer, A., Siedlarek, J.-P., 2017: "Community Leaders and the Preservation of Cultural Traits", *Journal of Economic Theory*, 168, 143–176

Przeworski, A. (2004): "The Last Instance: Are Institutions the Primary Cause of Economic Development?" *European Journal of Sociology*, 45(2), 165-188.

Schiavone, A. (1996): *La Storia Spezzata: Roma Antica e Occidente Moderno*, Bari, Laterza (new edition: Torino, Einaudi, 2020).

Tabellini, G. (2008a): "Institutions and Culture," Presidential address, *Journal of the European Economic Association*, 6(2-3), 255-94.

Tabellini, G. (2008b): "The Scope of Cooperation: Normes and Incentives," *Quarterly Journal of Economics*, 123(3), 905-950.

Tadei F. (2018), "The Long-Term Effects of Extractive Institutions: Evidence from Trade Policies in Colonial French Africa", Economic History of Developing Regions, 33, 3, 183-208.

Ticchi D., T. Verdier, and A. Vindigni (2013), ""Democracy, Dictatorship and the Cultural Transmission of Political Values," IZA Discussion Papers 7441.

Touré N. (2021), "Culture, Institutions and the Industrialization Process", *Journal of Economic Behavior & Organization,* 186, 481-503.

Treisman, D. (2017): "Democracy by Mistake," NBER WP 23944.

Valencia, F. (2021): "Historical Econometrics: Instrumental Variables and Regres-sion Discontinuity Designs," in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: ElsevierNorth Holland.

Verdier T. and Y. Zenou (2015), "The Role of Cultural Leaders in the Transmission of Preferences", *Economics Letters,* 136, C, 158-161.

Verdier T. and Y. Zenou (2018)," Cultural Leader and the Dynamics of Assimilation", *Journal of Economic Theory*, 175, 373-414.

Voth H.J. (2021), "Persistence – Myth and Mystery", in Bisin, A. and G.Federico (eds.), *Handbook of Historical Economics*, Amsterdam: Elsevier North Holland,.243-267.

Weibull, J. W. (1995),.*Evolutionary Game Theory.* MIT Press, Cambridge

Appendix A: Results on the dynamical system (6',7)

In this Appendix we study in some detail the dynamics of our economy. We study then the dynamics of $(\beta_t, q_t) \in [0,1]^2$. The fundamental institutional and cultural dynamics (6') and (7) are conveniently re-written to define the maps $f : [0,1]^2 \to [0,1]$ and $f : [0,1]^2 \to [0,1]$ as follows: the following:

$$\beta_{t+1} = f(\beta_t, q_t) := \begin{cases} \beta \text{ such that } p^{com}(\beta_t, q_{t+1}) = p(\beta, q_{t+1}) & \text{if it exists,} \\ \left[\begin{array}{ll} 1 & \text{if } P(\beta_t, q_{t+1}) > 0, \ \forall 0 \leq \beta \leq 1 \\ 0 & \text{if } P(\beta_t, , q_{t+1}) < 0 \ \forall 0 \leq \beta \leq 1 \end{array} \right. & \text{else} \end{cases}$$

$$q_{t+1} - q_t = g(\beta_t, q_t) := q_t(1 - q_t)S(\beta_{t+1}, q_{t+1})$$

We impose the following assumptions:

Assumption 1 *Utility functions are such that* $\mathbf{e}(\beta_t, q_t)$, $\mathbf{e}^{com}(\beta_t, q_t)$ *are continuous functions.*

Assumption 2 *For regularity we assume that* $p(\beta_t, q_t)$ *is monotonic in* β_t*.and that all maps,* $P(\beta, q)$, $S(\beta, q)$ *are smooth.*

We shall study the dynamical system in the continuous time limit, where the change in β_t and q_t between time t and $t + dt$ are, respectively, λdt and μdt, for $dt \to 0$.[51] [52]

$$\begin{aligned} \dot{\beta} &= \lambda \left[f(\beta, q) - \beta \right] \\ \dot{q} &= \mu g(\beta, q) \end{aligned} \tag{17}$$

given the initial conditions (β_0, q_0).

[51] As is well known, discrete time dynamics may generate complex dynamic behaviors that are difficult to characterize and go beyond the points we want to emphasize about the co-evolution between culture and institutions.

[52] This system is easily obtained in the following way. We assume that between time t and $t + dt$, an opportunity to change institutions arise with an instantaneous rate λdt Therefore the dynamics of β writes as:

$$\beta_{t+dt} = (1 - \lambda dt)\beta_t + \lambda dt f(\beta_t, q_{t+dt})$$

Similarly we may assume that between t and $t + dt$ a fraction μdt of individuals just before dying have an offspring socialized through cultural transmission. Then the dynamics of q writes as:

$$q_{t+dt} = (1 - \mu dt)q_t + \mu dt \left[q_t + g(\beta_{t+dt}, q_{t+dt}) \right]$$

Letting $dt \to 0$ provides immediately $\dot{\beta} = \lambda \left[f(\beta, q) - \beta \right]$ and $\dot{q} = \mu g(\beta, q)$.

The dynamics of β given q.

Lemma A. 1 *Under Assumptions 1-2, $f : [0,1]^2 \to [0,1]$ is a continuous function in $(\beta, q) \in [0,1]^2$.*

Proof. First of all note that when $p(\beta_{t+1}, q_{t+1}) = p^{com}(\beta_t, q_t)$ is not satisfied for any β_{t+1}, for some (q_t, q_{t+1}), the assumption that $p(\beta, q)$ is monotonic implies that β_{t+1} is $= 0$ or $= 1$, depending on the sign of $p^{com}(\beta_{t+1}, q_{t+1}) - p(\beta_t, q_t)$. In the continuous time limit $q_{t+1} = q_t = q$ and hence, in this case, trivially, f maps continuously $(\beta, q) \in [0,1]^2$ into $\{0\}$.

Consider equation (β), again. We show that β_{t+1} is a continuous function of β_t, q_t, q_{t+1} when $p(\beta_{t+1}, q_{t+1}) = p^{com}(\beta_t, q_t)$ is satisfied. To this end note that the assumed monotonicity in β of $p(\beta, q)$ implies that, when $p(\beta_{t+1}, q_{t+1}) = p^{com}(\beta_t, q_t)$ is satisfied, we can write $\beta_{t+1} = p^{-1}(p, q_t, q_{t+1})$ and hence $\beta_{t+1} = p^{-1}(p^{com}(\beta_t, q_t), q_t, q_{t+1})$, a continuous function. Again, in the continuous time limit $q_{t+1} = q_t$ and hence we can construct a continuous function $f : [0,1]^2 \to \mathbb{R}$ such that $\dot{\beta}_t = f(\beta_t, q_t)$.

Finally, it is straightforward to see that as $p^{com}(\beta_{t+1}, q_{t+1}) - p(\beta_t, q_t)$ crosses 0 $\beta_{t+1} = p^{-1}(p^{com}(\beta_t, q_t), q_t, q_{t+1})$ converges continuously to 0 or 1 depending on the direction of the crossing so as to preserve continuity. ■

Let the $\beta : [0,1] \to [0,1]$ map $q \in [0,1]$ into the stationary states of f; that is, $\beta : [0,1] \to [0,1]$ satisfies

$$0 = f(\beta, q), \text{ for any } \beta \in \beta(q)$$

Lemma A. 2 *Under Assumptions 1-2, the map $\beta : [0,1] \to [0,1]$ is an non empty and compact valued upper-hemi-continuous correspondence with connected components.*

Proof. The proof is a direct consequence of the continuity of f proved in Lemma A.1. ■

Let $P(\beta, q) := p^{com}(\beta, q) - p(\beta, q)$. We consider only the regular case in which $P(\beta, q) \neq 0$ at the vertices of $[0,1]^2$, leaving the simple but tedious analysis of the singular cases to the reader. Also, we say that q is a regular point of $\beta \in \beta(q)$ if any stationary stationary state $\beta \in \beta(q)$ satisfies that property that $\frac{\partial P(\beta,q)}{\partial \beta} \neq 0$; that is if $p(\beta, q)$ and $p^{com}(\beta, q)$ intersect transversally. The characterization of $\beta : [0,1] \to [0,1]$ depends crucially on the topological properties of the zeros of $P(\beta, q)$. Let $\pi : [0,1] \to [0,1]$ map q into the stationary states β such that $P(\beta, q) = 0$; that is, the map π satisfies $P(\pi(q), q) = 0$.

Proposition A. 1 *Under Assumptions 1-2, the dynamics of β as a function of $q \in [0,1]$ has the following properties,*

1. *$P(0,q) > 0$, $P(1,q) < 0$, for any $q \in [0,1]$, and $p(\beta,q)$ is increasing; or $P(0,q) < 0$, $P(1,q) > 0$, for any $q \in [0,1]$, and $p(\beta,q)$ is decreasing. For any given regular $q \in [0,1]$ there exist an odd number of regular stationary states $\beta \in \pi(q)$; furthermore $\beta = 0,1$ are also stationary states for given $q \in [0,1]$. The stability properties of the regular stationary states alternate with the smallest and the larger being always locally stable; the boundaries $\beta = 0,1$ are locally unstable for all $q \in [0,1]$.*

2. *$P(0,q) < 0$, $P(1,q) > 0$, for any $q \in [0,1]$, and $p(\beta,q)$ is increasing; or $P(0,q) > 0$, $P(1,q) < 0$, for any $q \in [0,1]$, and $p(\beta,q)$ is decreasing. For any given $q \in [0,1]$ there exist an odd number of regular stationary states $\beta \in \pi(q)$; furthermore $\beta = 0,1$ are also stationary states for given $q \in [0,1]$. The stability properties of the regular stationary states alternate with the smallest and the larger being always locally unstable; the boundaries $\beta = 0,1$ are locally stable.*

3. *$P(0,q) < 0$, $P(1,q) < 0$, for any $q \in [0,1]$. For any given $q \in [0,1]$ there exist either none or an even number of regular stationary states $\beta \in \pi(q)$; furthermore $\beta = 0$ is also a stationary state for given $q \in [0,1]$. The stability properties of the regular stationary states alternate with the smallest always locally unstable; the boundary $\beta = 0$ is locally stable.*

4. *$P(0,q) > 0$, $P(1,q) > 0$, for any $q \in [0,1]$. For any given $q \in [0,1]$ there exist either none or an even number of regular stationary states $\beta \in \pi(q)$; furthermore $\beta = 1$ is also a stationary state for given $q \in [0,1]$. The stability properties of the regular stationary states alternate with the smallest always locally stable; the boundary $\beta = 1$ is locally stable.*

5. *$P(0,q)$ and/or $P(1,q)$ change sign with $q \in [0,1]$. The characterization obtained above then can be repeated for each sub-interval of $[0,1]$ in which the Brouwer degree of the manifold $\pi(q)$ is invariant (see the proof). We leave the tedious categorization of all possible cases to the reader.*

The dynamics of q given β.

Lemma A. 3 *Under Assumptions 1-2, $g : [0,1]^2 \to [0,1]$ is a continuous function in $(\beta,q) \in [0,1]^2$.*

Proof. The proof is an immediate consequence of the continuity of $D(\beta_t, q_t)$, imposed in Assumption 3. ■

Let the $q : [0,1] \to [0,1]$ map $\beta \in [0,1]$ into the stationary states of g; that is, $q : [0,1] \to [0,1]$ satisfies

$$0 = g(\beta, q), \text{ for any } q \in q(\beta)$$

Lemma A. 4 *Under Assumptions 1-2, the map* $q : [0, 1] \to [0, 1]$ *is an non empty and compact valued upper-hemi-continuous correspondence with connected components. It contains* $q(\beta) = 0$ *and* $q(\beta) = 1$*, for any* $0 \leq \beta \leq 1$.

Proof. The proof is a direct consequence of the continuity of g, proved in Lemma A.1, and of the fact that $g(0, \beta) = g(1, \beta) = 1$, for any $0 \leq \beta \leq 1$. ■

As for the dynamics of culture given institutions, consider the regular case in which $S(\beta, q) \neq 0$ at the vertices of $[0, 1]^2$, leaving the simple but tedious analysis of the singular cases to the reader. We say that β is a regular point of $q \in q(\beta)$ if any stationary stationary state $q \in q(\beta)$ satisfies that property that $\frac{\partial S(\beta,q)}{\partial q} \neq 0$; that is if $d^i(\beta, q)$ and $d^j(1 - \beta, 1 - q)$ intersect transversally. The characterization of $q : [0, 1] \to [0, 1]$ depends crucially on the topological properties of the zeros of $S(\beta, q)$. Let $\sigma : [0, 1] \to [0, 1]$ map β into the stationary states q such that $S(\beta, q) = 0$; that is, the map σ satisfies $S(\sigma(\beta), \beta) = 0$.

Proposition A. 2 *Under Assumptions 1-2, the dynamics of* q *as a function of* $\beta \in [0, 1]$ *has the following properties,*

1. *$S(\beta, 0) < 0$, $S(\beta, 1) > 0$, for any $\beta \in [0, 1]$. For any given regular $\beta \in [0, 1]$ there exist an odd number of regular stationary states $q \in \sigma(\beta)$. By Lemma 4, $q = 0, 1$ are also stationary states for given $\beta \in [0, 1]$. The stability properties of the regular stationary states alternate starting with the $q = 0$ being stable and ending with $q = 1$ also being stable. If the dynamics supports a unique interior stationary state q^*, then it in unstable.*

2. *$S(\beta, 0) > 0$, $S(\beta, 1) < 0$, for any $q \in [0, 1]$. For any given regular $\beta \in [0, 1]$ there exist an odd number of regular stationary states $q \in \sigma(\beta)$. By Lemma 4, $q = 0, 1$ are also stationary states for given $\beta \in [0, 1]$. The stability properties of the regular stationary states alternate starting with the $q = 0$ being unstable and ending with $q = 1$ also being unstable. If the dynamics supports a unique interior stationary state q^*, then it in stable.*

3. *$S(\beta, 0) < 0$, $S(\beta, 1) < 0$, for any $\beta \in [0, 1]$. For any given regular $\beta \in [0, 1]$ there exist either none or an even number of regular stationary states $q \in \sigma(\beta)$. By Lemma 4, $q = 0, 1$ are also stationary states for given $\beta \in [0, 1]$. The stability properties of the regular stationary states alternate starting with the $q = 0$ being stable and ending with $q = 1$ being unstable.*

4. *$S(\beta, 0) > 0$, $S(\beta, 1) > 0$, for any $\beta \in [0, 1]$. For any given regular $\beta \in [0, 1]$ there exist either none or an even number of regular stationary states $q \in \sigma(\beta)$. By Lemma 4, $q = 0, 1$ are also stationary states for given $\beta \in [0, 1]$. The stability properties of the regular stationary states alternate starting with the $q = 0$ being unstable and ending with $q = 1$ being stable.*

5. *$S(\beta, 0)$ and/or $S(\beta, 1)$ change sign with $\beta \in [0, 1]$. The characterization obtained above then can be repeated for each sub-interval of $[0, 1]$ in which the Brouwer degree of the manifold*

$\sigma(\beta)$ is invariant (see the proof). We leave the tedious categorization of all possible cases to the reader.

Proof. Under Assumptions 1-2, $S(\beta, q)$ is smooth and (β, q) lie in the compact set $[0,1]^2$. $\sigma(\beta)$ is a dimension-1 smooth manifold with boundary , by a general version Implicit Function Theorem; see e.g. Milnor (1965), lemma 4, p. 13. The statement is then proved, closely along the lines of the proof of proposition A.1, using the full characterization of dimension-1 manifolds and Brouwer degree theory, thinking of $S(\beta, q)$ as a homothopy function varying β. We leave the details to the reader. ■

The joint dynamics of (β, q).

The dynamical system (6',7), even under Assumptions 1-2, is impossible to study in general. . We can however show that at least one stationary state always exists and characterize sufficient conditions for the existence of an interior stationary state. To this end we re-state here more formally Proposition 5 in the text.

Proposition A. 3 *Under Assumptions 1-2 the dynamical system (6',7) has at least one stationary state. Furthermore, if the Brouwer degree of both $\pi(q)$ and $\sigma(\beta)$ is ± 1, the dynamical system has at least one interior stationary state.*

Proof. The proof of the existence of a stationary state is a direct consequence of the characterization of $\beta(q)$ and $\sigma(\beta)$ in Lemmata A.2, A.4.

The proof of the existence of an interior stationary state under the Brouwer degree conditions is a consequence of the *Jordan curve theorem*, which we state in the following for completeness:[53]

A curve J in $\mathbb{R}^2$ which is the image of an injective continuous map of a circle into $\mathbb{R}^2$ has two components (an "inside" and "outside"), with J the boundary of each.

Figure A.1 represents a *Jordan curve J* on the plane.

[53] The theorem is a standard result in algebraic topology; see Hatcher (2002) p. 169 for a proof.

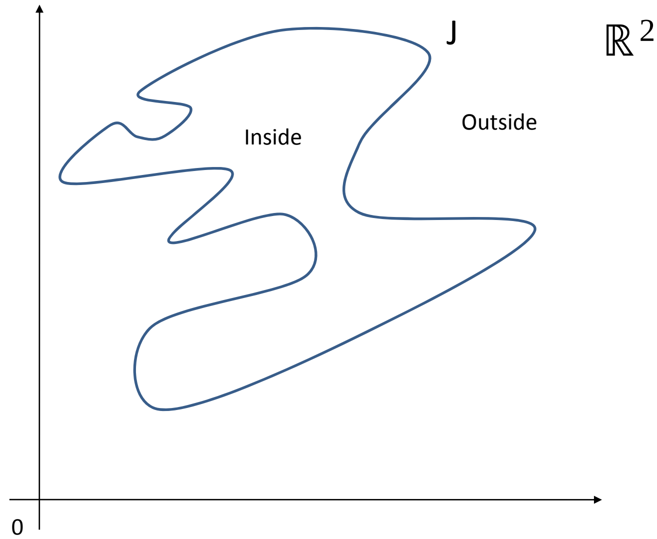


Figure A.1: Jordan Curve J in the (non negative) plane

Consider the compact space $[0,1]^2 \subset \mathbb{R}^2$, in which (β, q) lay. By Lemma 4, the map $q(\beta)$ contains the boundaries $q = 0, 1$ as well as the map $\sigma(\beta)$ which, in the case its Brouwer degree is ± 1, is homeomorphic to the compact interval $[0, 1]$. The map $\pi(q)$ is also homeomorphic to the compact interval $[0, 1]$ in the case its Brouwer degree is ± 1.

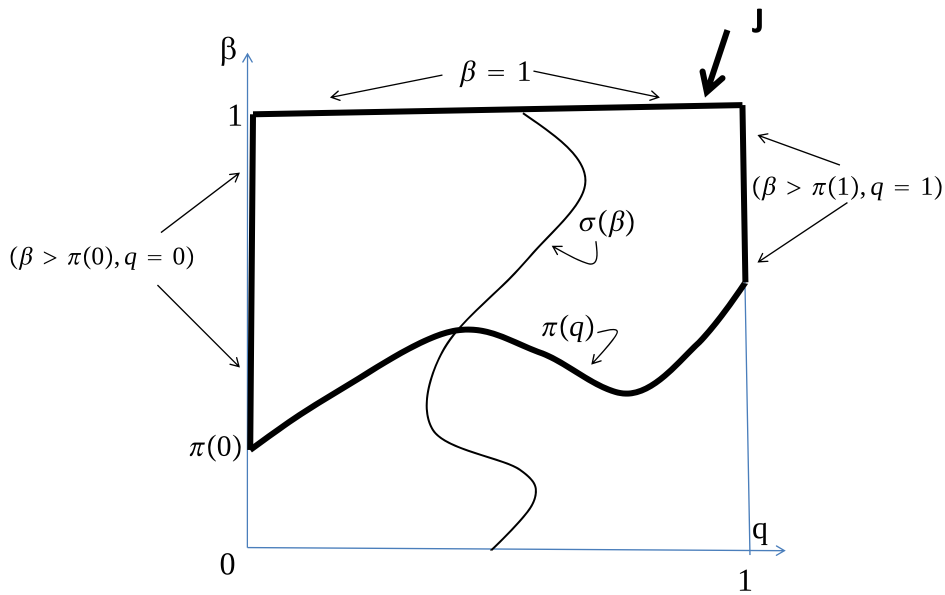


Figure A.2: Jordan Curve J in $[0,1]^2$ as constructed in the proof

We can therefore construct a Jordan curve J composed of $\pi(q)$, $(\beta > \pi(0), q = 0)$, $(\beta > \pi(1), q = 1)$, $\beta = 1$. Since $\sigma(\beta)$ connects the $\beta = 1$ and $\beta = 0$ its has a component inside and one outside the curve J. Furthermore, $0 < \sigma(\beta) < 1$, by construction. The *Jordan curve theorem* then guarantees that $\pi(q)$ and $\sigma(\beta)$ cross in the interior of $[0, 1]^2$; see Figure A.2 for a graphical representation of the construction. ■

Note that Proposition A.1 and A.2 provide conditions, respectively on $P(\beta, q)$ and $S(\beta, q)$, guaranteeing that the Brouwer degree of $\pi(q)$ and $\sigma(\beta)$ is ± 1. Also, the analysis leading to Proposition A.3 can be extended to dynamical systems in which the Brower degrees of $\pi(q)$ and $\sigma(\beta)$ are not invariant.

5.1 Further characterization of the Joint Dynamics

5.1.1 Complementarity and substitution between institutions and culture,cycles and oscillations

We provide here the proof of the propositions in the text which characterize the complex dynamics of culture and institutions.

Proof of **Proposition 5.** Suppose that conditions (18) are satisfied at an interior steady state

(β^*, q^*) of the system (17):

$$\frac{\partial P(\beta^*, q^*)}{\partial \beta}, \frac{\partial S(\beta^*, q^*)}{\partial q} < 0 \tag{18}$$

The linearized local dynamics around the interior steady state (β^*, q^*) can then easily be obtained by

$$\begin{pmatrix} \dot{\beta} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} \lambda \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right]_{(\beta^*,q^*)} & \lambda \left[\frac{\frac{\partial P}{\partial q}}{p_\beta}\right]_{(\beta^*,q^*)} \\ \mu q^*(1-q^*) \left[\frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)} & \mu q^*(1-q^*) \left[\frac{\partial S}{\partial q}\right]_{(\beta^*,q^*)} \end{pmatrix} \begin{pmatrix} \beta \\ q \end{pmatrix} \tag{19}$$

Recalling 1-2 and our normalization $p_\beta > 0$, then with enough regularity of the policy functions p^{com} and p, there exists a connected neighborhood of (β^*, q^*) such that the trace $T = \lambda \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right] + \mu q(1-q) \left[\frac{\partial S}{\partial q}\right]$ is negative and therefore does not change sign on that domain. The Bendixson Negative Criterion precludes then, in this case, the existence of local periodic orbits or limit cycles around (β^*, q^*) in that domain.

Note that when (18) are globally satisfied for all $(\beta, q) \in [0,1] \times [0,1]$,.it is not possible to get globally periodic orbits and limit cycles for dynamical system (17). Indeed given that in the simple connected domain $D = [0,1] \times [0,1]$.the sign of the trace $T = \lambda \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right] + \mu q(1-q) \left[\frac{\partial S}{\partial q}\right]$ is always strictly negative, the Bendixson Negative Criterion again precludes the existence of periodic orbits of (17) in this domain. ■

Proof of **Proposition 6.** Consider first an interior steady state (β^*, q^*) of (17) that is locally stable. Given our normalization $p_\beta > 0$ and the linearized system (19), we have the standard Hessian conditions

$$\begin{aligned} \frac{\partial P(\beta^*, q^*)}{\partial \beta} &< 0, \frac{\partial S(\beta^*, q^*)}{\partial q} < 0 \\ \left[\frac{\partial P}{\partial \beta} \cdot \frac{\partial S}{\partial q} - \frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)} &> 0 \end{aligned}$$

These local stability conditions ensure that the trace $T < 0$ and that the determinant $\Delta > 0$. Dampened oscillations (a stable spiral steady state equilibrium) require $T^2 < 4\Delta$. This last condition writes as :

$$\left[\lambda \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right]_{(\beta^*,q^*)} + \mu q^*(1-q^*) \left[\frac{\partial S}{\partial q}\right]_{(\beta^*,q^*)}\right]^2 < \frac{4\lambda\mu q^*(1-q^*)}{p_\beta(\beta^*,q^*)} \left[\frac{\partial P}{\partial \beta} \cdot \frac{\partial S}{\partial q} - \frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)};$$

or, after manipulations,

$$\left[\lambda \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right]_{(\beta^*,q^*)} - \mu q^*(1-q^*) \left[\frac{\partial S}{\partial q}\right]_{(\beta^*,q^*)}\right]^2 < -\frac{4\lambda\mu q^*(1-q^*)}{p_\beta(\beta^*,q^*)} \left[\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)}, \tag{20}$$

Given our normalization $p_\beta > 0$, when institutions and culture are dynamic complements at (β^*, q^*), we have $\frac{\partial P(\beta^*,q^*)}{\partial q}$ and $\frac{\partial S(\beta^*,q^*)}{\partial \beta}$ have the same sign. Hence $\left[\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right] > 0$ and the right hand side of inequality (20) is negative. Given that the left hand side is positive, it follows then that (20) cannot be satisfied and there are no dampening oscillations in cultural and institutional change when institutions and culture are dynamic complements at (β^*, q^*). ■

Proof of **the existence of dampened oscillations when culture and institutions are dynamic substitutes.** Conversely, assume now that culture and institutions are dynamic substitutes at the interior locally stable steady state (β^*, q^*). This implies that

$$\left[\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)} < 0$$

and dampening oscillations occur when (20) is satisfied. In this case, non-monotonic dynamics in culture and institutions obtain when

$$\frac{\lambda}{\mu}\left[\left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right]_{(\beta^*,q^*)} - \frac{\mu}{\lambda} q^*(1-q^*)\left[\frac{\partial S}{\partial q}\right]_{(\beta^*,q^*)}\right]^2 < \frac{4q^*(1-q^*)}{p_\beta(\beta^*,q^*)}\left[-\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)} \tag{21}$$

Using the local stability conditions for the Hessian at (β^*, q^*),

$$\begin{aligned} \left[\frac{\frac{\partial P}{\partial \beta}}{p_\beta}\right]_{(\beta^*,q^*)} &= -a < 0 \\ -q^*(1-q^*)\left[\frac{\partial S}{\partial q}\right]_{(\beta^*,q^*)} &= b > 0 \\ \frac{4q^*(1-q^*)}{p_\beta(\beta^*,q^*)}\left[-\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta}\right]_{(\beta^*,q^*)} &= M > 0. \end{aligned}$$

Denoting $x = \mu/\lambda$, the relative rate of change between culture and institutions, condition (21), can be written as

$$(-a + bx)^2 < Mx. \tag{22}$$

Simple examination of this condition reveals that (22) is satisfied when $x \in (x_- \, ; x_+)$, with

$$x_\pm = \frac{(2ab + M) \pm \sqrt{(2ab+M)^2 - 4(ab)^2}}{2b^2} > 0.$$

As a consequence, non-monotonic dynamics of institutions and culture around the locally stable steady state (β^*, q^*) are obtained when institutions and culture are dynamic substitutes and the relative rate of change between culture and institutions is neither too low, neither too high. ■

Proof of **Propositions 7.** Consider the cultural multiplier m_{SS} at a locally stable interior steady state (β^*, q^*). Recall the required normalizations:

$$p_\beta = \frac{\partial p(\beta^*, q^*, \gamma)}{\partial \beta} > 0, \; p_\gamma^{com} - p_\gamma = P_\gamma = \frac{\partial P(\beta^*, q^*, \gamma)}{\partial \gamma} > 0 \tag{23}$$

The comparative statics on (β^*, q^*) on the parameter are then easily obtained by differentiation of

$$\begin{aligned} P(\beta^*, q^*, \gamma) &= 0 \\ S(\beta^*, q^*) &= 0 \end{aligned} \tag{24}$$

one gets

$$\begin{aligned} \frac{d\beta^*}{d\gamma} &= \frac{-\frac{\partial S}{\partial q} P_\gamma}{\frac{\partial P}{\partial \beta}\frac{\partial S}{\partial q} - \frac{\partial P}{\partial q}\frac{\partial S}{\partial \beta}} \\ \frac{dq^*}{d\gamma} &= \frac{\frac{\partial S}{\partial \beta} P_\gamma}{\frac{\partial P}{\partial \beta}\frac{\partial S}{\partial q} - \frac{\partial P}{\partial q}\frac{\partial S}{\partial \beta}} \end{aligned} \tag{25}$$

Consider now the impact of a change in γ on institutional change, fixing q to its pre-shock value. Differentiating the first equation in (24),

$$\left(\frac{d\beta^*}{d\gamma}\right)_{q=q^*} = \frac{P_\gamma}{-\frac{\partial P}{\partial \beta}} > 0.$$

the stability condition for (β^*, q^*) require $\frac{\partial P}{\partial \beta} < 0$, $\frac{\partial S}{\partial q} < 0$ and $\Delta = \frac{\partial P}{\partial \beta}\frac{\partial S}{\partial q} - \frac{\partial P}{\partial q}\frac{\partial S}{\partial \beta} > 0$. Coupled with condition (23), this implies that the *cultural multiplier* on institutional change m, at (β^*, q^*), $m = \left(\frac{d\beta^*}{d\gamma}\right) / \left(\frac{d\beta^*}{d\gamma}\right)_{q=q^*} - 1$, is positive if and only if,

$$\frac{\partial P}{\partial q} \cdot \frac{\partial S}{\partial \beta} > 0$$

which is the condition for complementarity of the institutional and cultural dynamics. ■

Proof of **Propositions 8.** Consider the cultural multiplier m_{DD} on institutional change, from initial condition (β_0, q_0) in the basin of attraction B of a stationary state (β^*, q^*). . In this case the full dynamics of culture and institutions from (β_0, q_0) converges, by construction, to (β^*, q^*). In particular, institutions converge to $\beta^* = \beta(q^*)$ and $P(\beta(q^*), q^*) = 0$. In the counterfactual case in which the cultural composition of society had remained fixed, the dynamics of institutions would have converged to $\beta(q_0)$ and $P(\beta(q_0), q_0) = 0$. The cultural multiplier m_{DD} writes as

$$m_{DD} = \frac{\beta(q^*)}{\beta(q_0)} - 1.$$

Assume first that $q_0 < q^*$. Because institutions and culture are global dynamic complements, we know from proposition 6 that there are no dampened oscillations and therefore q_t monotonically increases from q_0 to q^*. Moreover, for all (β, q) in the basin of attraction B of (β^*, q^*), we should have $P_\beta(\beta, q) < 0$ as β converges to the stable manifold $\beta(q)$ for q inside the projection of B on the space of $q \in [0, 1]$.

Consider then first the case where $\frac{\partial P}{\partial q} > 0$, we get

$$P(\beta(q^*), q_0) < P(\beta(q^*), q^*) = 0 = P(\beta(q_0), q_0)$$

we get immediately that $\beta^* = \beta(q^*) > \beta(q_0)$ and $m_{DD} > 0$.

Similarly, when $\frac{\partial P}{\partial q} < 0$,

$$P(\beta(q^*), q_0) > P(\beta(q^*), q^*) = 0 = P(\beta(q_0), q_0)$$

and immediately $\beta^* = \beta(q^*) < \beta(q_0)$ and $m_{DD} < 0$. Consequently whatever the sign of $\frac{\partial P}{\partial q}$, m_{DD} has the same sign as $\frac{\partial P}{\partial q}$.

The case $q_0 > q^*$ can be handled by a similar argument, and one can easily see that m_{DD} should have the opposite sign as $\frac{\partial P}{\partial q}$ in such a case.

In conclusion m_{DD} has the same sign as $[q^* - q_0] \cdot \frac{\partial P}{\partial q}$. ■

5.1.2 Decomposition of the Cultural Multiplier on an aggregate variable $A(p, q, a^1(p), a^2(p))$

The *cultural multiplier* governs the effects of the interaction between culture and institutions on any aggregate economic variable of interest, e.g., per capita income, public good provision, or any other measure of economic activity. Let $A(p, q, a^1(p), a^2(p))$ formally denote the economic aggregate. A *cultural multiplier on A* can then be defined as

$$m_A = \frac{dA}{d\gamma} / \left(\frac{dA}{d\gamma}\right)_{q=q^*} - 1.$$

Noting from (25) that

$$\frac{dq^*}{d\gamma} = \frac{d\beta^*}{d\gamma} \cdot \frac{\frac{\partial S}{\partial \beta}}{-\frac{\partial S}{\partial q}}$$

we get

$$\frac{dA}{d\gamma} = \left\{ \underbrace{\left[A_p + \left(A_{a^1} a_p^1 + A_{a^2} a_p^2\right)\right] p_\beta}_{\text{direct effect}} + \underbrace{\left[A_q + \left[A_p + \left(A_{a^1} a_p^1 + A_{a^2} a_p^2\right)\right] p_q\right] \frac{\frac{\partial S}{\partial \beta}}{-\frac{\partial S}{\partial q}}}_{\text{indirect effect}} \right\} \frac{d\beta^*}{d\gamma}$$

The effect of γ on institutions will come from a direct effect as well as an indirect one. The direct effect in turn will be composed of two terms: a direct effect of the policy change induced by an institutional change p_β on the aggregate variable A (i.e., the term A_p), and the impact of changes in private actions $a^1(p)$ and $a^2(p)$ as induced also by the policy change p_β), the term $\left(A_{a^1}a^1_p + A_{a^2}a^2_p\right)p_\beta$). The indirect effect of cultural evolution will come from the compositional effect of changing the cultural group sizes (A_q), plus again the change in policy and private actions $\left[A_p + \left(A_{a^1}a^1_p + A_{a^2}a^2_p\right)\right]p_q$ which such a cultural compositional change induces.

Furthermore,

$$\left(\frac{dA}{d\gamma}\right)_{q=q^*} = \left[A_p + \left(A_{a^1}a^1_p + A_{a^2}a^2_p\right)\right]p_\beta \cdot \left(\frac{d\beta^*}{d\gamma}\right)_{q=q^*}.$$

Recalling that the cultural multiplier on institutions is $m = \left[\left(\frac{d\beta^*}{d\gamma}\right)/\left(\frac{d\beta^*}{d\gamma}\right)_{q=q^*} - 1\right]$,

$$m_A = \frac{dA}{d\gamma}/\left(\frac{dA}{d\gamma}\right)_{q=q^*} - 1 = m + K\cdot(1+m)$$

with

$$K = \frac{A_q + \left[A_p + \left(A_{a^1}a^1_p + A_{a^2}a^2_p\right)\right]p_q}{\left[A_p + \left(A_{a^1}a^1_p + A_{a^2}a^2_p\right)\right]p_\beta}\frac{\frac{\partial S}{\partial \beta}}{-\frac{\partial S}{\partial q}}$$

and hence m_A, can be expressed as a function of the institutional cultural multiplier m in terms of two components:

The first term is the cultural multiplier itself m. This reflects how the direct pass-through multiplier effect of institutions β on the aggregate variable $A(.)$ (through the impact of β on the equilibrium policy p, and individual behaviors). The second term $K\cdot(1+m)$ represents another multiplier effect on $A(.)$ that is triggered by the impact of institutional change on the cultural dynamics q, which in turn also affect the aggregate variable $A(.)$ through population effects. More precisely, this term is proportional to the relative degree of institutional change under joint evolution $(1+m) = \left(\frac{d\beta^*}{d\gamma}\right)/\left(\frac{d\beta^*}{d\gamma}\right)_{q=q^*}$ with a coefficient of proportionality K that reflects the sensitivity of cultural dynamics to institutions, as well as how the aggregate variable $A(.)$ depends on cultural change through a cultural group compositional effect and a culturally induced policy shift (ie. the effect of q on $p(\beta,q,\gamma)$). Depending on the sign of K, this second effect may either magnify or mitigate the direct pass-through cultural multiplier of institutional change on the variable A.

5.1.3 Cultural Dynamics in the Bisin and Verdier Model (2001)

Cultural transmission is modeled as the result of *direct vertical* (parental) socialization and *horizontal/oblique socialization* in society at large. *Direct vertical* socialization to the parent's trait $i \in I = \{1,2\}$ occurs with probability d^i. If a child from a family with trait i is not directly socialized, which occurs with probability $1-d^i$, he/she is *horizontally/obliquely* socialized by picking the trait of a role model chosen randomly in the population inside the political group (i.e., he/she picks trait i with probability q^i and trait $i' \neq i$ with probability $q^{i'}$).

If we let P^{ii} (resp. $P^{ii'}$) denote the probability that a child, in (a family in) group $i \in I$ is socialized to trait i (resp. i', we obtain:

$$P^{ii} = d^i + (1-d^i)q^i, \quad P^{ii'} = (1-d^i)q^{i'}$$

Let $V^{ii'}(\beta, q)$ denote the utility to a cultural trait i parent of a type i' child. It depends on the institutional set-up and the cultural distribution the child will face when he/she will make his/her economic decision $a^{i'}$.[54] Let $C(d^i)$ denote socialization costs. Direct socialization, for any $i \in I = \{1, 2\}$, is then the solution to the following parental socialization problem:

$$\max_{d^i \in [0,1]} -C(d^i) + \sum_{i' \in I} P^{ii'} V^{ii'}(\beta, q)$$
$$s.t. \;\; P^{ii} = d^i + (1-d^i)q^i, \quad P^{ii'} = (1-d^i)q^{i'}$$

As usual in this literature, define $\Delta V^i(\beta, q) = V^{ii}(\beta, q) - V^{ii'}(\beta, q)$ as the *cultural intolerance* of group i. It follows that direct socialization, $d^i(\beta, q)$ with some notational abuse is determined by the first order conditions:

$$\begin{aligned} C'(d^1) &= (1-q)\Delta V^1(\beta, q) \\ C'(d^2) &= (1-q)\Delta V^2(\beta, q) \end{aligned}$$

Turning again to the explicit notation for time t, the dynamics of q_t is straightforwardly determined by

$$q_{t+1} - q_t = q_t(1-q_t)S(\beta_{t+1}, q_{t+1})$$

with $S(\beta, q) = d^1(\beta, q) - d^2(\beta, q)$.

It is convenient to impose for regularity the following assumption (we do so in the examples) of separability of preference structures and quadratic costs of socialization:

Assumption 3 *$u^i\left(a^i, p; \mathbf{a}, q^i\right) = v^i\left(a^i, p\right) + H^i(p; \mathbf{a}, q^i)$, and $C(d^i) = \frac{1}{2}\left(d^i\right)^2$ for type $i = 1, 2$.*

Under (3), the cultural replicator dynamics (in continuous time) for fixed institutions β become

$$\dot{q_t} = q_t(1-q_t)S(\beta, q_t)$$

with $S(\beta, q)$ rewritten as:

$$S(\beta, q) = \left[\Delta V^1(p(\beta, q)) \cdot (1-q)\right] - \left[\Delta V^2(p(\beta, q)) \cdot q)\right]$$

[54] In extensive notation $V^{ii'}(\beta, q) = u^i\left(a^{i'}(\beta, q), p(\beta, q); a(\beta, q), q\right)$.

with

$$\begin{aligned}\Delta V^1(p) &= v^1(a^1(p),p) - v^1(a^2(p),p) \\ \Delta V^2(p) &= v^2(a^2\left(p\right),p) - v^2(a^1(p),p)\end{aligned}$$

Any interior stationary state q^* is obtained as a solution to:

$$\frac{\Delta V^1(p\left(\beta,q\right))}{\Delta V^2(p(\beta,q))} = \frac{q}{1-q}. \tag{26}$$

This equation may have many solutions characterizing the cultural steady state manifold $q = q^*(\beta)$. One may however provide sufficient conditions for the existence of a unique convergent cultura$l$ steady state. Specifically:

i) Assume that $\Delta V^1(p\left(\beta,0\right)) > 0$, $\Delta V^2(p(\beta,1)) > 0$ and the function $\phi\left(\beta,q\right) = \log\left[\frac{\Delta V^1(p(\beta,q))}{\Delta V^2(p(\beta,q)))}\right]$ is such that $\phi_q'\left(\beta,q\right) < 4$ for all $(\beta,q) \in [0,1]^2$, then (26) defines a unique solution $q^*(\beta)$ for every value of β and the cultural dynamics tend to that interior stationary state $q^*(\beta)$, whereby q increases when $q < q^*(\beta)$ and decreases instead when $q > q^*(\beta)$.

To see that, notice that at a point q^* satisfying $S(\beta,q^*) = 0$ or equivalently (26), one has

$$\left.\frac{\partial S}{\partial q}\right|_{q^*} = g'\cdot\left[\frac{d\Delta V^1}{dp}(1-q^*)\left.\frac{\partial p}{\partial q}\right|_{q^*} - \Delta V^1\right] - g'\cdot\left[\frac{d\Delta V^2}{dp}q^*\left.\frac{\partial p}{\partial q}\right|_{q^*} + \Delta V^2\right]$$

with $g' = g'(\Delta V^1(p(\beta,q^*))\cdot(1-q^*)) = g'(\Delta V^2(p(\beta,q^*))\cdot q^*))$. This is rewritten as

$$\left.\frac{\partial S}{\partial q}\right|_{q^*} = g'\left[q^*\Delta V^2\left(p^*\right)\left[\frac{1}{\Delta V^1}\frac{d\Delta V^1}{dp} - \frac{1}{\Delta V^2}\frac{d\Delta V^2}{dp}\right]\left.\frac{\partial p}{\partial q}\right|_{q^*} - \left(\Delta V^1(p^*) + \Delta V^2(p^*)\right)\right] < 0$$

Now, the condition $\phi_q'\left(\beta,q\right) < 4$ implies

$$\left[\frac{1}{\Delta V^1}\frac{d\Delta V^1}{dp} - \frac{1}{\Delta V^2}\frac{d\Delta V^2}{dp}\right]\cdot\frac{\partial p}{\partial q} < 4$$

This implies

$$\begin{aligned}\left.\frac{\partial S}{\partial q}\right|_{q^*} &< g'\left[4q^*\Delta V^2\left(p^*\right) - \left(\Delta V^1(p^*) + \Delta V^2(p^*)\right)\right] \\ &= \Delta V^2\left(p^*\right)g'\left[4q^* - 1 - \frac{\Delta V^1(p^*)}{\Delta V^2\left(p^*\right)}\right] \\ &= \Delta V^2\left(p^*\right)g'\left[4q^* - 1 - \frac{q^*}{1-q^*}\right] \\ &= \frac{\Delta V^2\left(p^*\right)g'}{1-q^*}\left[4q^*(1-q^*) - 1\right] < 0\end{aligned}$$

Hence for all point q^* satisfying $S(\beta, q^*) = 0$,one has $\left.\frac{\partial S}{\partial q}\right|_{q^*} < 0$. Given that $S(\beta, 0) = g(\Delta V^1(p(\beta, 0)) > 0$ and $S(\beta, 1) = -g(\Delta V^2(p(\beta, 1)) < 0$,this implies the uniqueness of $q^*(\beta)$ for every value of β, such that $S(\beta, q) > 0$ when $q < q^*(\beta)$ and $S(\beta, q) < 0$ when $q > q^*(\beta)$.

Note that the conditions for this result are in particular satisfied when $\Delta V^1(p)/\Delta V^2(p)$ is a decreasing (resp. increasing) function of the policy p and the equilibrium policy $p(\beta, q)$ is increasing (resp. decreasing) in q. Namely this happens when an increase of the frequency of a cultural trait induces a change of equilibrium policy which tends to reduce the relative marginal incentives (ie. paternalistic motive) of family transmission of that trait in the population. QED.

5.1.4 Linearized Joint Dynamics under Bisin and Verdier Model (2001)

We consider cultural transmission under the Bisin Verdier (20001) model (BV 2001 thereafter) with Assumption (3). Assumption (3) implies that $q(\beta) = \widehat{q}(p)$ with $p = p(\beta, q)$, with some notational abuse, where $\widehat{q}(p) \in [0, 1]$ is the unique solution of the following equation

$$\frac{\Delta V^1(p)}{\Delta V^2(p)} = \frac{q}{1-q}. \tag{27}$$

The separability of preference structures in Assumption (3) implies that the policy instrument p affects the optimal private actions, a^i, independently of the economy-level aggregates $\mathbf{a}$ and q. This in turn implies that cultural intolerances ΔV^i depend only on the equilibrium policy level p. As usual, we denote the partial derivative of a variable x on another variable y as $\partial x/\partial y = x_y$.

The linearized local dynamics around the interior steady state (β^*, q^*) of (17) can then easily be obtained by

$$\begin{pmatrix} \dot{\beta} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} \lambda \left[\frac{p_\beta^{com} - p_\beta}{p_\beta}\right]_{(\beta^*, q^*)} & \lambda \left[\frac{p_q^{com} - p_q}{p_\beta}\right]_{(\beta^*, q^*)} \\ -\mu G q^*(1-q^*) \cdot \widehat{q}_p \cdot p_\beta & \mu G q^*(1-q^*)\left[1 - \widehat{q}_p \cdot p_q\right] \end{pmatrix} \begin{pmatrix} \beta \\ q \end{pmatrix} \tag{28}$$

where $G = -(\Delta V^1(p(\beta^*, q^*)) + \Delta V^2(p(\beta^*, q^*)) < 0$.

The local stability of the interior steady state (β^*, q^*) of (17) is obtained under the standard Hessian conditions:

$$\begin{aligned} \left[\frac{p_\beta - p_\beta^{com}}{p_\beta}\right]_{(\beta^*, q^*)} &> 0 \\ 1 - \left[p_q \cdot \widehat{q}_p\right]_{(\beta^*, q^*)} &> 0 \\ \left[(1 - p_q \cdot \widehat{q}_p) \cdot \left[\frac{p_\beta - p_\beta^{com}}{p_\beta}\right] + \widehat{q}_p \cdot \left(p_q - p_q^{com}\right)\right]_{(\beta^*, q^*)} &> 0 \end{aligned} \tag{29}$$

The following Lemma characterizes the conditions for institutional and cultural dynamics to be complementary or substitute at a locally stable interior steady state.

Lemma A. 5 *With cultural evolution according to BV2001 and under Assumption 3, institutional and cultural* dynamics are complementary *at a locally stable interior steady state* (β^*, q^*) *if*

$$\frac{dP(\beta^*, q^*)}{dq} \text{ has the same sign as } \left[\frac{d\left(\frac{\Delta V^1(p)}{\Delta V^2(p)}\right)}{dp}\right]_{p(\beta^*,q^*)}; \tag{30}$$

they are instead substitute *if the signs are opposite.*

Proof. Institutional and cultural dynamics are complementary at (β^*, q^*) when

$$\frac{d\beta(q)}{dq} \text{ and } \frac{dq(\beta)}{d\beta} \text{ have the same sign.} \tag{31}$$

Differentiating,

$$\frac{d\beta(q)}{dq} = -\frac{\left(p_q - p_q^{com}\right)}{p_\beta - p_\beta^{com}}, \quad \frac{dq(\beta)}{d\beta} = \frac{\widehat{q}_p p_\beta}{1 - p_q \cdot \widehat{q}_p}.$$

Thus, condition (31) is equivalent to

$$\frac{d\beta(q)}{dq} \cdot \frac{dq(\beta)}{d\beta} \geq 0$$

or

$$-\frac{\left(p_q - p_q^{com}\right)}{p_\beta - p_\beta^{com}} \cdot \frac{\widehat{q}_p p_\beta}{1 - p_q \cdot \widehat{q}_p} \geq 0;$$

Given the Hessian conditions for local stability, (29), at an interior locally stable steady state (β^*, q^*), this condition is equivalent to

$$\left[\left(p_q^{com} - p_q\right) \cdot \widehat{q}_p\right]_{(\beta^*,q^*)} \geq 0.$$

Recalling that the cultural manifold $q(\beta)$ is obtained from

$$\frac{\Delta V^1(p)}{\Delta V^2(p)} = \frac{q}{1-q} \text{ and } p = p(\beta, q)$$

and that $P(\beta, q) := p^{com}(\beta, q) - p(\beta, q)$, differentiating,

$$\left[\left(p_q^{com} - p_q\right) \cdot \widehat{q}_p\right]_{(\beta^*,q^*)} = \left[P_q(\beta, q) \cdot \left[\frac{d\left(\frac{\Delta V^1(p)}{\Delta V^2(p)}\right)}{dp}\right]_{p(\beta,q)} (1-q)^2\right]_{(\beta^*,q^*)}.$$

Therefore, institutional and cultural *dynamics are complementary* at a locally stable interior steady state (β^*, q^*) when P_q and $\frac{d\left(\frac{\Delta V^1(p)}{\Delta V^2(p)}\right)}{dp}$ have the same sign at (β^*, q^*). Obviously they are dynamic substitute otherwise. ■

Appendix B: Extensions

We briefly discuss two main extensions to our analysis of the dynamics of culture and institutions in the text. First of all we consider the case in which $p(\beta, q_{t+1})$ is not necessarily monotonic, that is, when Assumption **??** is not satisfied. Second, we consider a general society in which political and cultural group are distinct, as in some of the example societies studied in Section 5.

Non-monotonic $p(\beta, q_{t+1})$.

Consider the case in which Assumption **??** is not imposed and hence $p(\beta, q_{t+1})$ can be non-monotonic. Then the dynamical system for β^i is characterized by the following implicit difference equation:

$$\beta^i_{t+1} = \begin{cases} \beta \text{ such that } p^{com}(\beta^i_t, q_{t+1}) = p(\beta, q_{t+1}) & \text{if it exists,} \\ \left[\begin{array}{ll} \arg\max p(\beta, q_{t+1}) & \text{if } p^{com}(\beta^I_t, q_{t+1}) > p(\beta, q_{t+1}),\ \forall 0 \le \beta \le 1 \\ \arg\min p(\beta, q_{t+1}) & \text{if } p^{com}(\beta^I_t, , q_{t+1}) < p(\beta, q_{t+1}),\ \forall 0 \le \beta \le 1 \end{array} \right. & \text{else} \end{cases} \tag{32}$$

In this case it is straightforward to show that the institutional dynamics might be underdetermined, that is, equation 32 might define an implicit map $(\beta^i_t, q_t) \to \beta^i_{t+1}$ which is multivalued in an open set of the domain. Furthermore, in this case, the dynamics of institutions can easily give rise to limit cycles. Consider for instance the example in Figure 1, with initial condition β_0, where the path $\beta_1 - > \beta_2 - > \beta_1$ constitutes such a limit cycle for a particular selection of the solutions to $P(\beta, q) = 0$.

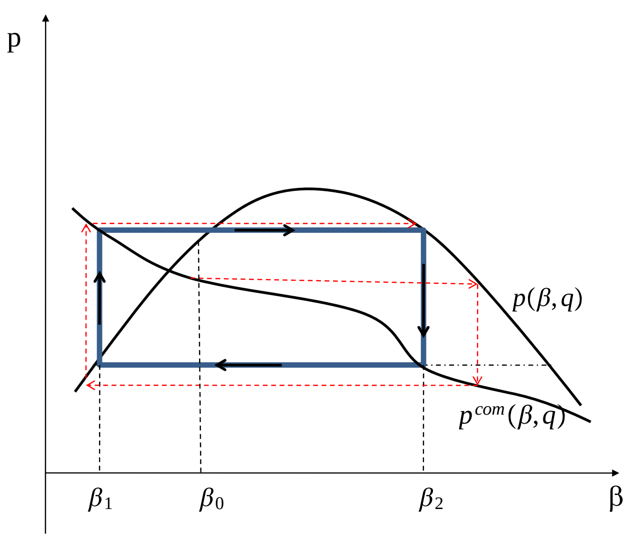


Figure A.3: Limit cycle (β_1, β_2) for $p(\beta,q)$ non-monotonic

Distinct political and cultural groups.

Consider a general society where political and cultural groups are distinct. We briefly indicate here how the concepts, Assumptions, and Propositions in the text are extended to this society. Let $i \in I$ index the political groups and $j \in J$ the cultural groups. Let a^{ij} denote the action of agents of subgroup (i, j) and $a = \{a^{ij}\}_{i,j}$ the vector profile of actions. Let q^{ij} denote the distribution of the population by cultural group and by $q = \{q^{ij}\}_{i,j}$ the vector profile satisfying $\sum_{j \in J} q^{ij} = 1$, for $i \in I$. Let λ^i denote the fraction of agents in political group i. Utility functions are then written as $u^{ij}\left(a^{ij}, p; a, q\right)$. We continue to identify political institutions with the weights of the groups $i \in I$ in the social choice problem which determines economic policy, $\beta = \{\beta^i\}_i$ satisfying $\sum_{i \in I} \beta^i = 1$.

The *societal equilibrium given institutions* β *and cultural distribution* q is a tuple $\{a, p\}$ such that:

$$\begin{array}{l} p \in \arg\max_p \sum_i \beta^i \sum_j q^{ij} u^{ij}\left(a^{ij}, p; a, q\right) \\ a^{ij} \in \arg\max\ u^{ij}\left(a^{ij}, p; a, q\right) \quad i \in I,\ j \in J. \end{array} \tag{33}$$

The *societal commitment equilibrium given institutions* β *and cultural distribution* q is a tuple $\{a^{com}, p^{com}\}$ such that:

$$\begin{array}{ll} \{a^{com}, p^{com}\} & \in \arg\max \sum_i \beta^i \sum_j q^{ij} u^{ij}\left(a^{ij}, p; a, q\right) \\ & s.t. \quad a^{ij} \in \arg\max\ u^{ij}\left(a^{ij}, p; a, q\right), \quad i \in I,\ j \in J \end{array} \tag{34}$$

Restricting to dychotomous groups, that is $I = \{1, 2\}$ and $J = \{a, b\}$ the *societal equilibrium*, the *societal commitment equilibrium*, and the *societal optimum* can be denoted, respectively:

$$[a(\beta, q), p(\beta, q)]\,; \quad [a^{com}(\beta, q), p^{com}(\beta, q)]\,; \quad \left[a^{eff}(\beta, q), p^{eff}(\beta, q)\right]$$

Assumption B. 1 *Utility functions are sufficiently regular so that* $a(\beta, q),\ p(\beta, q),\ a^{com}(\beta, q),\ p^{com}(\beta, q)$ *are continuous functions.*

Assumption B. 2 *Utility functions are sufficiently regular so that* $p(\beta, q)$ *is monotonic in* β.

Adding an index t to denote time, institutions evolve as a solution to the following design problem:

$$\max_{\beta_{t+1}} \sum_{i \in I} \beta_t^i \sum_{j \in J} q_{t+1}^{ij} u^{ij}\left(a^{ij}(\beta_{t+1}, q_{t+1}), p(\beta_{t+1}, q_{t+1}); a(\beta_{t+1}, q_{t+1})\right) \tag{35}$$

Proposition B. 1 *Under Assumption 1-2, and given* (q_t, q_{t+1})*, the dynamics of institutions* β_t^i*,* $i \in I$*, is governed by the following implicit difference equation:*

$$\beta_{t+1}^i = \left\{ \begin{array}{ll} \beta^i \text{ such that } p^{com}(\beta, q_{t+1}) = p(\beta_t, q_{t+1}) & \text{if it exists,} \\ \left[\begin{array}{ll} 1 & \text{if } p^{com}(\beta_t, q_{t+1}) > p(\beta, q_{t+1}),\ \forall 0 \le \beta^i \le 1 \\ 0 & \text{if } p^{com}(\beta_t, q_{t+1}) < p(\beta, q_{t+1}),\ \forall 0 \le \beta^i \le 1 \end{array} \right. & \text{else} \end{array} \right. \tag{36}$$

It is convenient to define $P(\beta, q) := p^{com}(\beta, q) - p(\beta, q)$.

Proposition B. 2 *Under Assumption 1-2, for any given q, the dynamics of institutions governed by (36) have at least one stationary state. An* interior stationary states β^* *obtains as a solution to* $P(\beta, q) = 0$. *The* boundary stationary state $\beta^i = 1$ *obtains when* $P(\beta, q) \mid_{\beta^i=1} > 0$; *while the* boundary stationary state $\beta^i = 0$ *obtains when* $P(\beta, q) \mid_{\beta^i=0} < 0$.

Proposition B. 3 *Under Assumption 1-2, for any given q, in the continuous time limit, the dynamics governed by (36) satisfies the following properties:*

if $P(\beta, q) > 0$ *for any* $\beta^i \in [0, 1]$, *then* $\beta^i = 1$ *is a globally stable stationary state.*

if $P(\beta, q) < 0$ *for any* $\beta^i \in [0, 1]$, *then* $\beta^i = 0$ *is a globally stable stationary state;*

any boundary stationary state is always locally stable;

if an interior stationary state β^* *exists, it is locally stable if* $\frac{\partial P(\beta^*, q)}{\partial \beta^i} < 0$.

As for cultural transmission, we assume for simplicity that political groups are culturally perfectly segregated, so that the reference population for an agent in subgroup (i, j) is the subgroup itself. Fixing a political group $i \in I$, cultural evolution can then be written as a general equation system of the form:

$$q_{t+1}^{ij} - q_t^{ij} = q^{ij}(1 - q_t^{ij}) S^{ij}(\beta_{t+1}, q_{t+1}^{ij}) \text{ for } i \in I, \ j \in J$$

Specializing to the BV (2001) model of cultural transmission, direct vertical socialization to the parent's trait, say $j \in J$, occurs with probability d^{ij}; $P_t^{i,jj}$ (resp. $P_t^{i,jj'}$) denote the probability that a child, in (a family in) political group $i \in I$ with trait j is socialized to trait j (resp. j') at t; $V^{i,jj}(\beta_{t+1}, q_{t+1})$ (resp. $V^{i,jj'}(\beta_{t+1}, q_{t+1})$) denotes the utility to a cultural trait j parent in political group i of a type j (resp. j') child. Cultural transmission implies:

$$\begin{aligned} P_t^{i,jj} &= d^{ij} + (1 - d^{ij}) q_t^{ij} \\ P^{i,jj'} &= (1 - d^{ij})(1 - q_t^{ij}) \end{aligned}$$

$$V^{i,jj}(\beta_{t+1}, q_{t+1}) = u^{ij}\left(a^{ij}(\beta_{t+1}, q_{t+1}), p(\beta_{t+1}, q_{t+1}); a(\beta_{t+1}, q_{t+1}), q_{t+1}\right) \quad (37)$$

$$V^{i,j\neq j}(\beta_{t+1}, q_{t+1}) = u^{ij}\left(a^{ij'}(\beta_{t+1}, q_{t+1}), p(\beta_{t+1}, q_{t+1}); a(\beta_{t+1}, q_{t+1}), q_{t+1}\right) \quad (38)$$

Let $C(d^{ij})$ denote socialization costs. Direct socialization is then the solution to the following parental socialization problem:

$$\max_{d^{ij} \in [0,1]} -C(d^{ij}) + P_t^{i,jj} V^{i,jj}(\beta_{t+1}, q_{t+1}) + P_t^{i,jj'} V^{i,jj'}(\beta_{t+1}, q_{t+1}), \text{ s. t. 1) }.$$

Calling $\Delta V^{ij}(\beta_{t+1}, q_{t+1}) = V^{i,jj}(\beta_{t+1}, q_{t+1}) - V^{i,j\neq j}(\beta_{t+1}, q_{t+1})$, the *cultural intolerance* of trait j in political group i, it follows that the direct socialization, with some notational abuse, has the form:

$$d^{ij} = d^{ij}(q_t, \Delta V^{ij}(\beta_{t+1}, q_{t+1})) = d^{ij}(\beta, q), \ i \in I, \ j \in J \tag{39}$$

Assumption B. 3 *Utility and socialization cost functions are sufficiently regular so that $d^{ij} = d^{ij}(\beta, q)$ is continuous.*

Proposition B. 4 *Under Assumption 3, and given β_{t+1}, the dynamics of culture q_t^{ij} is governed by the following difference equation:*

$$q_{t+1}^{ij} - q_t^{ij} = q_t^{ij}(1 - q_t^{ij})\left(d^{ij} - d^{ij'}\right). \tag{40}$$

evaluated at $d^{ij} = d^{ij}(q_t, \Delta V^{ij}(\beta_{t+1}, q_{t+1}))$ satisfying (39).

It is convenient to define $S^{ij}(\beta, q) := d^{ij}(\beta, q) - d^{ij'}(\beta, q)$.

Proposition B. 5 *Under Assumption 3, for any given β, the dynamics of institutions governed by (40) have at least the two boundary stationary states, $q^{ij} = 0$ and $q^{ij} = 1$. An* interior stationary states $0 < q^{ij*} < 1$ *obtains as a solution to $S^{ij}(\beta, q) = 0$.*

Proposition B. 6 *Under Assumption 3, for any given β, in the continuous time limit, the dynamics governed by (40) satisfies the following properties:*

if $S^{ij}(\beta, q) > 0$ for any $q^{ij} \in [0, 1]$, then q_t^{ij} converges to $q^{ij} = 1$ from any initial condition $q_0^{ij} > 0$;

if $S^{ij}(\beta, q) < 0$ for any $q^{ij} \in [0, 1]$, then q_t^{ij} converges to $q^{ij} = 0$ from any initial condition $q_0^{ij} < 1$;

if $S^{ij}(\beta, 1) > 0$, then $q^{ij} = 1$ is locally stable ;

if $S^{ij}(\beta, 0) < 0$, then $q^{ij} = 0$ is locally stable;

if an interior stationary state q^{ij*} exists, and $\frac{\partial S^{ij}(\beta, q^{ij*})}{\partial q^{ij}} < 0$, it is locally stable.

Consider the following assumption:

Assumption B. 4 *Socialization costs are quadratic:*

$$C(d^{ij}) = \frac{1}{2}\left(d^{ij}\right)^2.$$

We then obtain:

Corollary B. 1 *Under Assumption 4,*

$$S^{ij}(\beta, q) = \Delta V^{ij}(\beta, q)q^{ij'} - \Delta V^{ij'}(\beta, q)q^{ij},$$

and hence interior steady states are characterized by solutions to:

$$\frac{\Delta V^{ij}(\beta, q)}{\Delta V^{ij'}(\beta, 1-q)} = \frac{q^{ij}}{q^{ij'}} \tag{41}$$

Under Assumptions 1-3, the joint dynamics of institutions and culture is governed by the system (36,40), which we report here for convenience:

$$\begin{aligned} \beta^i_{t+1} &= \begin{cases} \beta^i \text{ such that } p^{com}(\beta, q_{t+1}) - p(\beta_t, q_{t+1}) & \text{if it exists,} \\ \left[\begin{array}{ll} 1 & \text{if } p^{com}(\beta_t, q_{t+1}) > p(\beta, q_{t+1}), \ \forall 0 \leq \beta^i \leq 1 \\ 0 & \text{if } p^{com}(\beta_t, , q_{t+1}) < p(\beta, q_{t+1}), \ \forall 0 \leq \beta^i \leq 1 \end{array} \right. & \text{else} \end{cases} \\ q^{ij}_{t+1} - q^{ij}_t &= q^{ij}(1 - q^{ij}_t)\left(d^{ij} - d^{ij'}\right), \text{ with } d^{ij} = d^{ij}(q_t, \Delta V^{ij}(\beta_{t+1}, q_{t+1})). \end{aligned}$$

Proposition B. 7 *Under Assumptions 1-3 the dynamical system (36,40) has at least one stationary state. Furthermore, if both the institutional and the cultural dynamics display an interior stationary state, respectively, for all $0 \leq q^{ij} \leq$ and all $0 \leq \beta \leq 1$, then the dynamical system (36,40) has at least one interior stationary state.*

The slippery slope argument

Define more specifically the Nash equilibrium behavioral vector of society for a given policy level p as $\mathbf{a}(p, q) = \left[a^i(p, q)\right]_{i=1,2}$ [55]. The policy maker objective function associated to such equilibrium behavioral response associated to a policy p rewrite as

$$\overline{W}(\beta_t, p, q) = W(\beta_t, \mathbf{a}(p, q), p, q)$$

for the current period. [56]

[55]For simplicity, we assume that conditions hold to ensure the existence of a unique such Nash equilibrium behavioral vector $\mathbf{a}(p, q)$. Formally it is the fixed point of the following correspondence T which associates to each behavioral vector $\mathbf{a} = \left[a^i\right]_{i=1,2}$, the behavioral vector $T(\mathbf{a}) = \left[\widetilde{a}^i(p, q, \mathbf{a})\right]_{i=1,2}$ such that

$$\widetilde{a}^i(p, q, \mathbf{a}) \in \arg\max_a \ u^i(a, p; \mathbf{a}, q)$$

[56]We consider the case where $\beta^0(\beta_{t+1}) \in (0, 1)$, is an interior solution. The argument can be appropriately accommodated when $\beta^0(\beta_{t+1}) = 0, 1$.

For current institutions β_t, the "one-step" forward looking institutional design is obtained from the perfect Nash equilibrium $\left(\beta^1(\beta_t), \widetilde{\beta}^0(\beta_t)\right)$ of the following two-period game. In the first stage, initial institutions β_t design institutions β_{t+1},. In the second stage, the institutions β_{t+1} in turn design next period institutions β'_{t+2} in a myopic way. $\left(\beta^1(\beta_t), \widetilde{\beta}^0(\beta_t)\right)$ should then satisfy:

$$\beta^1(\beta_t) \in \arg\max_{\beta'} \overline{W}\left(\beta_t, p\left(\beta', q\right), q\right) + \frac{\delta}{1-\delta}\overline{W}\left(\beta_t, p\left(\beta^0\left(\beta'\right), q\right), q\right)$$

where $\beta^0(\beta)$ is the optimal myopic institutions β_{t+2} designed by a given second stage institution player β (ie.

$$\beta^0(\beta) \in \arg\max_{\beta'} \frac{1}{1-\delta}\overline{W}\left(\beta, p\left(\beta', q\right), q\right)$$

and

$$\widetilde{\beta}^0(\beta_t) = \beta^0\left(\beta^1(\beta_t)\right)$$

One can solve the game by backward induction. In stage 2, the institutional "player" β_{t+1} is myopic and given that social welfare evaluated at weights β_{t+1} is highest under policy $p^{com}(\beta_{t+1}, q)$, the optimal institutional choice β^0_{t+2} is designed to induce the choice $p^{com}(\beta^1_{t+1}, q)$; that is $p(\beta^0_{t+2}, q) = p^{com}(\beta^1_{t+1}, q)$ or $\beta^0_{t+2} = \beta^0(\beta_{t+1})$. Consequently in stage 1, the optimal choice of institutions by the current institutional framework β_t, is given by

$$\beta^1_{t+1} \in \arg\max_{\beta'} \overline{W}\left(\beta_t, p\left(\beta', q\right), q\right) + \frac{\delta}{1-\delta}\overline{W}\left(\beta_t, p^{com}(\beta', q), q\right)$$

Define the function

$$\Psi(\beta_t, \beta, q) == \overline{W}(\beta_t, p(\beta, q), q) + \frac{\delta}{1-\delta}\overline{W}(\beta_t, p^{com}(\beta, q), q)$$

and assume that the function $\Psi(\beta_t, \beta, q)$ is smooth and strictly concave in β, so that a first order approach holds. Formally, we require that the policy functions $p(\beta, q)$ and $p^{com}(\beta, q)$ are smooth and that

$$\overline{W}_{pp}\left(\frac{\partial p}{\partial \beta}\right)^2 + \overline{W}_p \frac{\partial^2 p}{\partial \beta^2} + \frac{\delta}{1-\delta}\overline{W}^{com}_{pp}\left(\frac{\partial p^{com}}{\partial \beta}\right)^2 + \overline{W}_p \frac{\partial^2 p^{com}}{\partial \beta^2} < 0$$

This will be ensured when the function $\overline{W}(\beta_t, p, q) = W(\beta_t, \mathbf{a}(p, q), p, q)$ is a well defined smooth enough concave function of the policy level p.[57]

[57] This is ensured when:

$$\frac{\partial^2 W}{\partial p^2} < 0$$

and

$$\overline{W}_{pp}(\beta_t, p, q) = \sum_{i=1,2} \frac{\partial W}{\partial a^i} \cdot \frac{\partial^2 a^i}{\partial p^2} + 2\sum_{i=1,2} \frac{\partial^2 W}{\partial a^i \partial p}\frac{\partial ai}{\partial p} + \frac{\partial^2 W}{\partial p^2}$$

is sufficiently negative

An interior solution $\beta^1_{t+1} = \beta^1(\beta_t) \in (0,1)$ should then satisfy:

$$\Psi_\beta\left(\beta_t, \beta^1_{t+1}, q\right) = \overline{W}_p\left(\beta_t, p\left(\beta^1_{t+1}, q\right), q\right) \cdot \left.\frac{\partial p}{\partial \beta}\right|_{\beta^1_{t+1}} + \frac{\delta}{1-\delta}\overline{W}_p\left(\beta_t, p^{com}(\beta^1_{t+1}, q), q\right) \cdot \left.\frac{\partial p^{com}}{\partial \beta}\right|_{\beta^1_{t+1}} = 0$$

To illustrate the impact of the "slippery slope" argument, let us compare for a given current institutional player β_t the "one-step forward looking" institutional design $\beta^1(\beta_t)$ to the "myopic" choice $\beta^0(\beta_t)$. Suppose also that the myopic institutional design path $\beta_{t+1} = \beta^0(\beta_t)$ converges monotonically to the stable institutional steady state $\overline{\beta} \in (0,1)$ and assume that $p^{com}(\beta, q)$ is increasing in β. As we know, this steady state is characterized by $p\left(\overline{\beta}, q\right) = p^{com}(\overline{\beta}, q)$, $p(\beta, q)$ is also increasing in β and $p(\beta, q) \geq p^{com}(\beta, q)$ if and only if $\beta \leq \overline{\beta}$.

i) Note first that the steady state $\overline{\beta}$ is also a steady state of the "one step forward-looking" institutional dynamics design $\beta_{t+1} = \beta^1(\beta_t)$ (ie $\overline{\beta} = \beta^1_{t+1}\left(\overline{\beta}\right)$. Indeed taking into account the fact that $p\left(\overline{\beta}, q\right) = p^{com}(\overline{\beta}, q)$ and that $\overline{W}_p(\beta, p^{com}(\beta, q), q) = 0$, we have

$$\begin{aligned}
\Psi_\beta\left(\overline{\beta}, \overline{\beta}, q\right) &= \overline{W}_p\left(\overline{\beta}, p\left(\overline{\beta}, q\right), q\right) \cdot \left.\frac{\partial p}{\partial \beta}\right|_{\overline{\beta}} + \frac{\delta}{1-\delta}\overline{W}_p\left(\overline{\beta}, p^{com}(\overline{\beta}, q), q\right) \cdot \left.\frac{\partial p^{com}}{\partial \beta}\right|_{\overline{\beta}} \\
&= \overline{W}_p\left(\overline{\beta}, p^{com}(\overline{\beta}, q), q\right) \cdot \left[\left.\frac{\partial p}{\partial \beta}\right|_{\overline{\beta}} + \frac{\delta}{1-\delta}\left.\frac{\partial p^{com}}{\partial \beta}\right|_{\overline{\beta}}\right] = 0
\end{aligned}$$

ii) As well note that

$$\begin{aligned}
\Psi_\beta(\beta_t, \beta_t, q) &= \overline{W}_p(\beta_t, p(\beta_t, q), q) \cdot \left.\frac{\partial p}{\partial \beta}\right|_{\beta_t} + \frac{\delta}{1-\delta}\overline{W}_p(\beta_t, p^{com}(\beta_t, q), q) \cdot \left.\frac{\partial p^{com}}{\partial \beta}\right|_{\beta_t} \\
&= \overline{W}_p(\beta_t, p(\beta_t, q), q) \cdot \left.\frac{\partial p}{\partial \beta}\right|_{\beta_t}
\end{aligned}$$

The concavity of $\overline{W}$ in p and the fact that $\overline{W}$ reaches its maximum at $p^{com}(\beta, q)$, implies that $\Psi_\beta(\beta_t, \beta_t, q) \geq 0$ iff $p(\beta_t, q) \leq p^{com}(\beta_t, q)$ and $\Psi_\beta(\beta_t, \beta_t, q) \geq 0$ if and only if $\beta_t \leq \overline{\beta}$. The concavity of $\Psi(\beta_t, \beta, q)$ in β, implies therefore that $\beta_t \leq \beta^1(\beta_t)$ if and only if $\beta_t \leq \overline{\beta}$.

iii) Note finally that at point $\beta^0_{t+1} = \beta^0(\beta_t)$ by definition, one has $p^{com}(\beta^0_{t+1}, q) = p(\beta_t, q)$ and again noting that $\overline{W}_p(\beta_t, p^{com}(\beta_t, q), q) = 0$, one gets

$$\Psi_\beta\left(\beta_t, \beta^0(\beta_t), q\right) = \frac{\delta}{1-\delta}\overline{W}_p\left(\beta_t, p^{com}(\beta^0(\beta_t), q), q\right) \cdot \left.\frac{\partial p^{com}}{\partial \beta}\right|_{\beta^0(\beta_t)}$$

the sign of which depends on the sign of $\overline{W}_p\left(\beta_t, p^{com}(\beta^0(\beta_t), q), q\right)$. Thus, with the concavity of $\overline{W}$ in p and the fact that $\overline{W}$ reaches its maximum at $p^{com}(\beta, q)$, one has $\Psi_\beta\left(\beta_t, \beta^0(\beta_t), q\right) \geq 0$ if and only if $p^{com}(\beta^0(\beta_t), q) \leq p^{com}(\beta_t, q)$ or $\beta^0(\beta_t) \leq \beta_t$. The concavity of $\Psi(\beta_t, \beta, q)$ in β, implies that $\beta^0(\beta_t) \leq \beta^1(\beta_t)$ if and only if $\beta^0(\beta_t) \leq \beta_t$.

For all $\beta_t < \overline{\beta}$, we have $\beta_t < \beta^0(\beta_t)$ and consequently $\beta_t < \beta^1(\beta_t) < \beta^0(\beta_t)$ while for $\beta_t > \overline{\beta}$, we have $\beta_t > \beta^0(\beta_t)$ and consequently $\beta^0(\beta_t) < \beta^1(\beta_t) < \beta_t$. From this we can conclude that the "one-step forward looking" institutional change converges to the same steady state $\overline{\beta}$ as the myopic institutional change but at a reduced speed.

Appendix C: Assumptions on fundamentals

In this Appendix we translate Assumptions 1-2 into restrictions on fundamentals.

Sufficient conditions for the existence and monotonicity of the societal equilibrium $p(\beta, q)$.

Without loss of generality, restrict $p \in [0,1]$. The indirect utility function can de written as:

$$u^i(\mathbf{e}{,}q) = u^i(a^i, p; A, q);$$

where the individual private action $a^i \in [0,1]$ and A is an aggregate population level index $A = A(a^1, a^2, p, q)$.

Assume $u^i(.)$ is twice differentiable in $(a^i, p; A, q)$ and strictly concave in a^i (ie. $u^i_{11} < 0$). Assume also that the aggregator function $A(.)$ is differentiable in (a^1, a^2, p, q) and such that the image of $[0,1]^4$ by $A(.)$ is an interval $[A_{\min}; A_{\max}]$. Finally, assume the following boundary conditions:

$$u^i_1(0, p; A, q) \geq 0, \; u^i_1(1, p; A, q) \leq 0 \;\text{ for all } (p, A, q) \in [0,1] \times [A_{\min}; A_{\max}] \times [0,1].$$

These conditions and the fact that $u^i(.)$ is a strictly concave function in a^i ensure that the optimal individual behavior for a given value of p and A is characterized by a continuous function $a^i(p, A, q) \in [0,1]$ obtained from the First Order Condition:

$$u^i_1(a^i, p; A, q) = 0.$$

For given values of $p \in P$ and $q \in [0,1]$, a Nash equilibrium in private actions a^{1N}, a^{2N} and aggregate index $A^N(p,q)$ is characterized by the solution of the following system:

$$a^{iN} = a^i\left(p, A^N, q\right) \quad \text{for } \; i \in (1,2) \; \text{ and } A^N = A(a^{1N}, a^{2N}, p, q),$$

which in turn translates into the following condition for A^N:

$$A^N = A(a^1\left(p, A^N, q\right), \; a^2\left(p, A^N, q\right), p, q). \tag{42}$$

The following sufficient conditions ensure the existence of a unique Nash equilibrium in private actions $a^{1N}(p,q)$, $a^{2N}(p,q)$, $A^N(p,q)$:

$$\begin{aligned} 1 - \sum_{i=1,2} A'_i \frac{u^i_{13}}{-u^i_{11}} &> 0 \;\text{ for all } (a^1, a^{2,}, A, p, q) \\ A(a^1(p, A_{\min}, q), a^2(p, A_{\min}, q), q) &> A_{\min} \;\text{ for all } (p,q) \in [0,1]^2 \\ A(a^1(p, A_{\max}, q), a^2(p, A_{\max}, q), q) &< A_{\max} \;\text{ for all } (p,q) \in [0,1]^2. \end{aligned}$$

The first condition ensures that the function $\Gamma(x,p,q) = x - A(a^1(p,x,q), a^2(p,x,q), p, q)$ is increasing for all $(p,q) \in [0,1]^2$. The second and the third conditions ensure that $\Gamma(A_{\min}, p, q) < 0 < \Gamma(A_{\max}, p, q)$; Together these conditions ensure the existence of a unique value $A^N(p,q)$ satisfying (42) and thus correspondingly a unique Nash equilibrium profile $a^{1N}(p,q), a^{2N}(p,q)$.

Moreover, differentiating,

$$\begin{aligned} \frac{dA^N}{dp} &= \frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]} \\ \frac{da^{iN}}{dp} &= \frac{u^i_{12}}{-u^i_{11}} + \frac{u^i_{13}}{-u^i_{11}} \frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]}. \end{aligned}$$

The condition for an interior *societal equilibrium* $p(\beta, q)$ is obtained from the First Order Conditions of the policymaker,

$$\beta\ u^1_2(a^1, p, A, q) + (1-\beta)\ u^2_2\left(a^2, p, A, q\right) = 0.$$

After substitution of the Nash equibrium private actions $a^{1N}(p,q), a^{2N}(p,q), A^N(p,q)$, this condition can be written as

$$\Psi(p,q,\beta) = 0; \tag{43}$$

with

$$\Psi(p,q,\beta) = \beta\ u^1_2(a^{1N}(p,q), p, A^N(p,q), q) + (1-\beta)\ u^2_2\left(a^{2N}(p,q), p, A^N(p,q), q\right).$$

Moreover a corner societal equilibrium $p(\beta,q) = 0$ (resp. $p(\beta,q) = 1$) obtains when $\Psi(0,q,\beta) \leq 0$ (resp. $\Psi(1,q,\beta) \geq 1$).

A sufficient condition for the existence of a unique *societal equilibrium* $p(\beta,q)$ consists in the function $\Psi(p,q,\beta)$ being decreasing in p for all $q \in [0,1]$. Given the smoothness assumptions on the functions $u^i(.)$ and $A(.)$ this is satisfied when the following condition holds:

$$u^i_{12}\frac{da^i}{dp} + u^i_{22} + u^i_{23}\frac{dA}{dp} < 0 \ \text{ for all } i \in (1,2).$$

In turn, in terms of the fundamentals, this conditions becomes:

$$\frac{u^i_{12}}{-u^i_{22}}\left[\frac{u^i_{12}}{-u^i_{11}} + \frac{u^i_{13}}{-u^i_{11}} \frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]}\right] + \frac{u^i_{23}}{-u^i_{22}} \frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]} < 1;$$

or

$$\frac{\left(u^i_{12}\right)^2}{u^i_{22}u^i_{11}} + \left(\frac{u^i_{12}}{u^i_{22}}\frac{u^i_{13}}{u^i_{11}} + \frac{u^i_{23}}{(-u^i_{22})}\right)\frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]} < 1 \quad \text{for } i \in (1,2)\,,$$

with $A_p = \partial A/\partial p$, and $A_j = \partial A/\partial a^j$,. This condition is more likely to be satisfied when $\left|u^i_{11}\right|$ and $\left|u^i_{22}\right|$ are large enough.

To obtain a condition ensuring the monotonicity in β of the *societal equilibrium* $p(\beta, q)$, we differentiate (5.1.4), obtaining

$$\frac{\partial p}{\partial \beta} = \frac{u^1_2 - u^2_2}{-\Psi_p}.$$

Thus $p(\beta, q)$ is monotonic in β when $u^1_2 - u^2_2$ has a constant sign.

This condition simplifies if the preferences structure is characterized by some degree of separability:

$$\begin{aligned} u^i(a, p; A, q) &= u(a, p; A, q, \theta_i) \\ &= v\,(a, p, \theta_i) + H(p, A) \end{aligned}$$

and $\theta_1 > \theta_2$. Such preferences lead to

$$\begin{aligned} a^{1N} &= a(p, \theta_1) \\ a^{2N} &= a(p, \theta_2) \\ A^N &= A(a(p, \theta_1), a(p, \theta_2), p, q). \end{aligned}$$

A sufficient condition for the existence of a unique *societal equilibrium*, given that $u^j_{13} = 0$, is then

$$\frac{\left(v^i_{12}\right)^2}{\left[v^i_{22} + H_{pp}\right] v^i_{11}} + \left(\frac{H_{pA}}{-(v^i_{22} + H_{pp})}\right)\left[A_p + \sum_{j=1,2} A'_j \frac{v^j_{12}}{-v^j_{11}}\right] < 1 \quad \text{for } i \in (1,2)\,,$$

where $v^i_{kl} = v"_{kl}\,(a, p, \theta_i)\,.$ But

$$\begin{aligned} u^1_2 - u^2_2 &= u_2(a^{1N}, p, A^N, q, \theta_1) - u_2\left(a^{2N}, p, A^N, q, \theta_2\right) \\ &= v_2\left(a(p, \theta_1), p, \theta_1\right) - v_2\left(a(p, \theta_2), p, \theta_2\right). \end{aligned}$$

Thus a sufficient conditions for the monotonicity of the *societal equilibrium* $p(\beta, q)$ is that $v_p\left(a(p, \theta), p, \theta\right)$ is monotonic in θ; or, after manipulations, that $v_{ap}\frac{v_{a\theta}}{(-v_{aa})} + v_{p\theta}$ has a constant sign.

Consider as an example the following preference structure:

$$u(a, p, A, q, \theta) = (1 - p)a + \theta W(1 - a) + H(p, A);$$

with $W(.)$ a strictly increasing and concave function, $A = qa^1 + (1-q)a^2$, $H(p, A)$ concave in p. Then,

$$\begin{aligned} v_a &= (1-p) - \theta W'(1-a),\ v_p = -a \\ v_{ap} &= -1, v_{a\theta} = -W'(1-a) \\ -v_{aa} &= -\theta W"(1-a) \\ v_{p\theta} &= 0,\ v_{pp} = 0. \end{aligned}$$

The sufficient condition for a well defined *societal equilibrium* $p(\beta, q)$ can be written as

$$\frac{1}{H_{pp}\theta_i W"} + \left(\frac{H_{pA}}{-(H_{pp})}\right)\left[\sum_{j=1,2} q_j \frac{1}{\theta_j W"}\right] < 1 text\ for i = 1, 2.$$

When $H_{pA} > 0$, given that $-(H_{pp}) > 0$ and $\left[\sum_{j=1,2} q_j \frac{1}{\theta_j W"}\right] < 0$, this condition is satisfied when $\frac{1}{H_{pp}\theta_i W"} < 1$, which in turn holds when $1 < H_{pp}W"\theta_2$. This is satisfied when $H_{pp}W"$ is sufficiently large; that is, with enough concavity of W and H, respectively, in a and p.

When $H_{pA} < 0$, this sufficient condition can be rewritten as

$$\frac{1}{\theta_i} - H_{pA}\left[\sum_{j=1,2} q_j \frac{1}{\theta_j}\right] < H_{pp}W",$$

which again will be satisfied when H_{pA} is bounded from below on the relevant domain, $[0,1] \times [A_{\min}, A_{\max}]$ (i.e., $H_{pA} > -K$, with $K > 0$) and $H_{pp}W" > (1+K)/\theta_2$. This also is satisfied with enough concavity of W and H, respectively, in a and p.

Finally, the monotonicity of the *societal equilibrium* function $p(\beta, q)$ holds when $v_{ap}\frac{v_{a\theta}}{(-v_{aa})} + v_{p\theta}$ has a constant sign. But

$$v_{ap}\frac{v_{a\theta}}{(-v_{aa})} + v_{p\theta} = \frac{W'}{-\theta W"} > 0.$$

Thus the *societal equilibrium* function is monotonically increasing in β.

Sufficient conditions for the existence of the societal commitment equilibrium $p^{com}(\beta, q)$.

The *societal commitment equilibrium given institutions β and cultural distribution q* is obtained from the following maximization problem:

$$\begin{array}{l} \max\ \beta\, u^1\left(a^{1N}, p; A^N, q\right) + (1-\beta)\, u^2\left(a^{2^N}, p; A^N, q\right) \\ s.t.\ \ a^{iN} = a^{iN}(p, q) \quad \text{for} \quad i \in (1,2) \ \text{ and } A^N = A^N(p, q) \end{array} . \tag{44}$$

Let

$$\begin{aligned}\Omega\left(p,\beta,q\right) &= \beta\; u^1\left(a^{1N}\left(p,q\right),p;A^N(p,q),q\right)\\ &\quad +(1-\beta)\; u^2\left(a^{1N}\left(p,q\right),p;A^N(p,q),q\right).\end{aligned}$$

The First Order Condition for an interior *societal commitment equilibrium* $p^{com}(\beta,q)$ can be written as

$$\begin{aligned}\Omega_p\left(p,\beta,q\right) &= \beta\; u_2^1(a^{1N},p,A^N,q)+(1-\beta)\; u_2^2\left(a^{2N},p,A^N,q\right)\\ &\quad +\left(\beta\; u_3^1(a^{1N},p,A^N,q)+(1-\beta)\; u_3^2\left(a^{2N},p,A^N,q\right)\right)\frac{dA^N}{dp}\\ &= 0.\end{aligned}$$

A sufficient (strong) condition is then,

$$\Omega_{pp}(p,\beta,q)<0,\;\text{ for all } p\in[0,1]\,.$$

Differentiating,

$$\begin{aligned}\Omega_{pp}(p,\beta,q) &= \sum_{i=1,2}\beta^i(u_{12}^i\frac{da^{iN}}{dp}+u_{22}^i+u_{23}^i\frac{dA^N}{dp})\\ &\quad +\sum_{i=1,2}\beta^i\left[u_{13}^i\frac{da^{iN}}{dp}+u_{23}^i+u_{33}^i\right]\frac{dA^N}{dp}\\ &\quad +\sum_{i=1,2}\beta^i u_3^i\frac{d^2A^N}{dp^2}.\end{aligned}$$

Thus a sufficient condition for $\Omega_{pp}(p,\beta,q)<0$ is that, for $i=1,2$,

$$u_{12}^i\frac{da^{iN}}{dp}+u_{22}^i+u_{23}^i\frac{dA^N}{dp}+\left[u_{13}^i\frac{da^{iN}}{dp}+u_{23}^i+u_{33}^i\right]\frac{dA^N}{dp}+u_3^i\frac{d^2A^N}{dp^2}<0.$$

Recall

$$\frac{da^{iN}}{dp}=\frac{u_{12}^i}{-u_{11}^i}+\frac{u_{13}^i}{-u_{11}^i}\frac{A_p+\sum_{j=1,2}A_j'\frac{u_{12}^j}{-u_{11}^j}}{\left[1-\sum_{j=1,2}A_j'\frac{u_{13}^j}{-u_{11}^j}\right]}$$

and

$$\frac{dA^N}{dp}=\frac{A_p+\sum_{j=1,2}A_j'\frac{u_{12}^j}{-u_{11}^j}}{\left[1-\sum_{j=1,2}A_j'\frac{u_{13}^j}{-u_{11}^j}\right]}.$$

Tedious manipulations show then that a sufficient condition for $\Omega_{pp}(p,\beta,q)<0$ is that

$$D^i = \frac{\left(u^i_{12}\right)^2}{-u^i_{11}} + u^i_{22} + \left(2(\frac{u^i_{13}u^i_{12}}{-u^i_{11}} + u^i_{23}) + u^i_{33}\right) \frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]}$$
$$+\frac{\left(u^i_{13}\right)^2}{-u^i_{11}} \left[\frac{A_p + \sum_{j=1,2} A'_j \frac{u^j_{12}}{-u^j_{11}}}{\left[1 - \sum_{j=1,2} A'_j \frac{u^j_{13}}{-u^j_{11}}\right]}\right]^2 + u^i_3 \frac{d^2A^N}{dp^2}$$

is negative for $i=1,2$. Because of the term in d^2A^N/dp^2, this involves complicated conditions on the third derivatives of the indirect preference functions. When preferences are separable, of the form

$$\begin{aligned} u^i(a,p;A,q) &= u(a,p;A,q,\theta_i) \\ &= v\left(a,p,\theta_i\right) + H(p,A), \end{aligned}$$

the expression D^i simplifies somewhat:

$$\begin{aligned} D^i &= \frac{\left(v^i_{ap}\right)^2}{-v^i_{aa}} + v^i_{pp} + \left(2H_{pA} + H_{AA}\right)\left(A_p + \sum_{j=1,2} A'_j \frac{v^j_{ap}}{-v^j_{pp}}\right) \\ &\quad + H_A \frac{d^2A^N}{dp^2}. \end{aligned}$$

Therefore $\Omega\left(p,\beta,q\right)$ is strictly concave in p when $v\left(a,p,\theta_i\right)$ is sufficiently concave in (a,p).