Essays on Political Economy and Cultural Evolution

Avner Seror

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Ecole des hautes études en sciences sociales

Essais en Economie Politique et sur l'Evolution des Normes Culturelles

Ecole doctorale n°465

ECOLE DOCTORALE ECONOMIE PANTHEON SORBONNE

Spécialité ECONOMIE

COMPOSITION DU JURY :

M. BISIN Alberto
New York University, Rapporteur

M. TORVIK Ragnar
Dragvoll University, Rapporteur

M. ALDASHEV Gani
Université Libre de Bruxelles, Membre du jury

Mme. KRANTON Rachel
Duke University, Membre du jury

M. VERDIER Thierry
EHESS-PSE, Membre du jury

Soutenue par Avner **SEROR**
le 19 Mars 2018

Dirigée par **Thierry VERDIER**

L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES | PSL

Remerciements

I had the chance to be surrounded by extraordinarily insightful and wide-ranging intellectuals that allowed me to contribute to our understanding of culture and institutions during my years at the Paris School of Economics. The first essay of this dissertation studies child development and rearing practices. The second presents a theory on the co-evolution of cultural norms and economic institutions, while the last essay focuses on political institutions and digs into the issue of political competition.

Je tiens avant tout à remercier Thierry Verdier qui a été un directeur de thèse exceptionnel. Je lui suis reconnaissant pour son dévouement, ses conseils, ses encouragements et son sens de l'humour qui a toujours accompagné nos discussions. Je le remercie également d'avoir accordé une bourse de thèse au jeune étudiant de master que j'étais. Je n'avais postulé à aucune autre. L'un des articles de cette dissertation est un travail commun. Peut-être que ça pourrait marcher avec des distributions de Fréchet, avait-il dit. Je l'arrêtais alors, invoquant une sombre contrainte pesant sur notre stratégie naissante. Il acquiesça et la chose fut oubliée. Deux mois plus tard, les bases d'un modèle de vote probabiliste utilisant des Fréchets étaient mises sur paper. Et ça fonctionnait. . .plutôt bien.

I am grateful to the members of the jury, Gani Aldashev, Alberto Bisin, Rachel Kranton and Ragnar Torvik. I am particularly grateful to Gani Aldashev, who has been a member of the committee of my thesis and who has closely followed my work for the last three years. I am grateful to Alberto Bisin for accepting to review this dissertation and for inviting me at New York University. This short stay was the opportunity to initiate very fruitful discussions on the joint evolution of culture and institutions. I am also grateful to Ragnar Torvik, who accepted to review this dissertation. My Ph.D has been financed by the E.R.C grant n° 324004 that is gratefully acknowledge.

I am also grateful to Roland Bénabou, Jared Rubin and Laurence Iannaccone for their invaluable comments on my study of the evolution of religious preferences. I am grateful to Nathan Nunn, the editor of the Journal of Development Economics, and two anonymous referees for their insightful comments on this work as well. I would like to thank Ithzak Gilboa for having discussed at an early stage my work on the formation of preferences and intergenerational rearing. I am indebted as well to Laurent Bouton and Micael Castanheira for their comments and help on our common work with Thierry Verdier on multi-candidate elections.

Je remercie également Sultan Mehmood, Emeline Bezin et Clémentine Van Effenterre pour nos discussions, qui je l'espère donneront lieu à de futures contributions communes. J'ai une pensée particulière pour Clémentine, qui m'a permis de découvrir des travaux importants

en sociologie, en philosophie et en anthropologie. Je suis également reconnaissant à Emeline pour son aide tout au long de ma thèse. I am grateful to Sultan for sharing his ideas and impressing intuitions with me on the political economy of Pakistan. Merci à ceux qui ont su motiver mon travail et éveiller ma curiosité. Merci à mes amis Adrien, Hadrien, David, Arthur, Margaux, Simon, Lara, Cédric et Caro.

Je suis infiniment redevable à mes proches. Merci à mes deux frères Jarod et Jérémie. A ma soeur Myriam qui relit depuis des années déjà la plupart de mes écrits. Merci à son mari Mickael, merci à leur fils Méir qui m'a regardé avec attention écrire ces quelques lignes du haut de ses quelques jours. Merci à ma mère. Merci à mon père. A ma grand mère Elizabeth. Et merci à mon grand père, papi Bob, jusqu'à 120 ans.

Avner Seror,
29th December 2018,
Paris

A mes parents, et à mon grand père Yacov Bénisti,

Summary

Modern economics has accepted the challenge to understand matters that half a century lay firmly outside its realm: culture and institutions. This dissertation presents three essays on these issues.

The first essay of the dissertation presents presents a theory of child development and parental rearing practices. In the model, a benevolent parent seeks to transmit behavioral norms to her child by sending him signals that are imperfectly observed. The child can however increase the quality of the signals he receives by investing in acquiring cognitive skills. We establish that neither authoritarian nor permissive parenting styles are conducive to the accumulation of cognitive skills. Furthermore, since rearing interactions aim at transmitting behavioral norms, the child develops a capital of appreciation for particular cognitive skills. This cultural perspective to the issue of cognitive development provides an interpretation grid for various results established in the empirical literature on child development. Our approach also permits to identify the parental characteristics that are particularly conducive to child maltreatment. We predict that child maltreatment should be expected to emerge early in the rearing period and to be persistent, hence the need for early and repeated interventions for maltreated children. We also suggest that the use of television and video materials for rearing purposes can widen the existing disparities in the early acquisition of cognitive skills.

The second essay provides a unifying theory for the existence of religious prohibition against secular sciences or usury. In the model, a religious leader controls the religious doctrine and decides whether to discriminate some economic activities. The agents allocate their labor in one of two productive sectors and there exist monopoly producers of sector-specific technologies. One sector is potentially subject to prohibition because it aggregates the occupations that are complementary to scientific discoveries for example. Prohibition aims at altering the dynamics of religious preferences in a way that maximizes the rents of some religious leader. The central intuition of this chapter is that prohibition creates a cultural segmentation of the labor market. Furthermore, a culturally segmented labor market affects the dynamics of cultural norms because it differentiates cultural groups. In turn, such an "economic" distinction between cultural groups adds a salient dimension to parents' existing incentives to transmit their cultural norms intergenerationally. This is why prohibition can substantially increase the diffusion of religious preferences in the population, and can therefore substantially benefit religious leaders. This framework also yields interesting predictions relative to the degree of collusion between politics and religion and relative to the duration of economic prohibitions.

The purpose of the last essay is to provide a simple analytical framework on multi-candidate elections that is tractable enough to allow for a systematic analysis of the endogenous structure

of political competition in ways similar to the Industrial Organization models used to analyse endogenous market structures under economic competition. The versatility of our approach is demonstrated through several applications on classical topics in political economy: special interest politics, coalition formation in the legislature, and franchise extension in the democracy. Among the key contributions of the paper, we establish that both proportional and run-off systems should lead to higher party fragmentation than a plurality system. Furthermore, we are able to show that allowing high campaign investments leads to less fragmented polities and more rent extraction from well funded political parties.

Field: Economics
Key words: Political Economy; Cultural Evolution; Cultural Transmission; Child Development

Résumé

L'économie moderne a accepté le défi d'étudier des problématiques qui il y a encore un demi-siècle se trouvaient en dehors de son domaine d'expertise, comme l'évolution de la culture et des institutions. Cette dissertation présente trois essais sur ces questions.

Le premier essai présente une théorie sur le développement de l'enfant et les pratiques parentales. Dans le modèle, un parent cherche à transmettre des normes comportementales à son enfant en lui envoyant des signaux que l'enfant observe de manière imparfaite. L'enfant peut cependant augmenter la qualité des signaux qu'il reçoit en investissant dans l'acquisition de compétences cognitives. Nous établissons que les styles parentaux autoritaires ou permissifs décroissent l'accumulation de compétences cognitives. De plus, puisque les interactions entre parents et enfants visent à transmettre des normes comportementales, l'enfant développe un capital d'appréciation pour le développement de compétences cognitives. Notre perspective culturelle sur la question du développement cognitif fournit une grille d'interprétation utile pour divers résultats établis dans la littérature empirique sur le développement de l'enfant. Notre approche permet également d'identifier les caractéristiques parentales qui sont susceptibles de conduire à la maltraitance des enfants. Nous prédisons que la maltraitance des enfants est un comportement parental qui, si optimal, est persistant et apparaît tôt dans la période d'éducation, d'où la nécessité d'interventions précoces et répétées pour les enfants maltraités. Nous suggérons également que l'utilisation de la télévision et de vidéos à des fins d'éducation peut élargir les disparités existantes dans l'acquisition précoce de compétences cognitives.

Le second essai de cette dissertation présente une théorie sur l'existence de prohibitions religieuses contre les sciences séculières ou contre l'usure. Dans le modèle, un chef religieux contrôle la doctrine religieuse et décide si celle-ci doit être hostile à un ensemble d'activités économiques comme les activités scientifiques ou innovantes par exemple. Les prohibitions visent à altérer la trajectoire de l'évolution des normes religieuses de manière à maximiser les rentes captées par le chef religieux. Les agents allouent leur force de travail étant donné que certaines occupations peuvent être sujettes à l'hostilité de la doctrine religieuse. L'intuition centrale de cet essai c'est que les prohibitions créent une segmentation culturelle du marché du travail. En effet, alors que le coût des prohibitions pèse sur l'utilité des religieux et les amène à choisir des occupations non-prohibées, les non-religieux préfèrent choisir les occupations prohibées de sorte à bénéficier d'un premium sur leur salaire. Cette segmentation culturelle du marché du travail, en conséquence, crée une distinction entre les groupes culturels, et change la trajectoire de l'évolution des normes culturelles. En particulier, il est possible sous certaines conditions que les prohibitions économiques augmentent la taille du groupe religieux, auquel cas nous comprenons pourquoi il peut être optimal pour un chef religieux d'inclure

une dose de prohibitions économiques dans la doctrine religieuse. Cet essai permet également d'approcher la problématique de l'existence de collusion entre pouvoirs politiques et religieux, ainsi que de comprendre quels sont les facteurs sociaux affectant la durée des prohibitions économiques.

Le dernier essai de cette dissertation présente un cadre analytique simple pour l'étude d'élections impliquant plusieurs candidats. Le modèle est suffisamment simple pour permettre une analyse systématique de la structure endogène de la compétition politique, et est ainsi similaire aux théories d'Organisation Industrielle permettant l'étude de la structure des marchés et de la compétition économique. La souplesse de notre approche est démontrée à travers divers applications sur des sujets standards d'économie politique comme la politique de redistribution, la formation de coalition dans la législature, et l'extension de la franchise. Parmi les contributions centrales de ce papier, nous établissons que les systèmes proportionnels, ainsi que les systèmes majoritaires favorisent une plus grande fragmentation politique que les systèmes pluralistes. De plus, nous démontrons que l'autorisation d'investissements de campagne électorale importants crée une structure hétérogène du marché politique, qui mêle des partis "low cost" créent à partir de faibles financements ainsi que des partis "high cost" capturant plus de voix et plus de rentes.

Discipline : Sciences économiques
Mots-clés : Economie politique; évolution culturelle; transmission culturelle; développement de l'enfant.

Table of contents

Main Introduction

Modern economics has accepted the challenge to understand matters that half a century lay firmly outside its realm: culture and institutions. This dissertation presents three essays on these issues.

While scholars have identified important determinants of norms and institutions in recent years, there is still a deep remaining conundrum, as argued recently by Kranton (2016, p. 407): "why do divisions and norms resonate for human beings?" This interrogation motivated the first essay of this dissertation, which looks upstream of the relationship between culture and institutions and studies the formation of cultural identity through intergenerational interactions with caregivers. It has been recently argued by Bisin and Verdier (2015) that knowing the *origin* of either culture or institution loses much of its interest since it is the joint evolution of culture and institutions that determines economic growth and prosperity. In their own words, "the focus is moved from the cause (both culture and/or institution can have causal effects) to the process that determine the interaction" Bisin and Verdier (2015, p. 3). The second chapter of this dissertation studies the joint evolution of a religious identity and economic and political institutions. Finally, the regulation of political competition is a major characteristic of political institutions. Yet the process of selection of political rulers is more than often modeled by political economists as a two-party contest for power. This simplification overshadows the study of the socio-economic determinants of political competition, and therefore constrains to a large extent the analysis of the determination of political institutions. Filling this gap is the main motivation of the third essay of this dissertation, which is a common work with Pr. Thierry Verdier.

In this introduction, I will first review the literature on culture and institutions that is relevant for this thesis. A brief outline of the analysis of the three chapters of the dissertation will then be provided.

Culture and institutions: an overview

Few will now underestimate the importance of culture and institutions in the process of development. An influential body of research has demonstrated that economic and institutional backwardness are a byproduct of history, since they are observed in economies that experienced despotic governments and extractive institutions (North (1981); Acemoglu, Johnson, and Robinson (2001); Tabellini (2008)). As a simple and well known illustration, Putnam, Leonardi, and Nanetti (1993) link the contemporary variations in the efficiency of regional governments in Italy to cultural values, which origin could traced back to the prevalence of free cities during the middle ages. Relatedly, Grosfeld, Sakalli, and Zhuravskaya (2016) provide empirical support that Italian cities that achieved self-government in the Middle Ages have a higher level of civic capital today than similar cities in the same area that did not, and suggest that the Middle-Age experience of self-government fostered the inter-generational transmission of self-efficacy beliefs. As another well known example, Nunn and Wantchekon (2011) have shown that the transatlantic and Indian Ocean slave trades created a culture of distrust within Africa that still persists today.[1] Finally, Greif (1993, 2006a,b); Greif and Tabellini (2010, 2017) have studied the interaction between cultural norms, the family - an institution of primer importance in human societies - and the institutions governing trade relations. The preceding contributions demonstrate that the nuclear family in medieval times helped establish and grow corporations in Europe. By contrast, extended kinship groups allowed for trust-based trade relationships. Only corporations however fostered beliefs and norms that justify and support self-governance, the rule of law, the legitimacy of majority rule, respect for minority rights, individualism and trust among nonkin (Greif (2006b)).

1. Other illustrations can be found in the studies of Bisin and Verdier (2015) and Alesina and Giuliano (2015).

In turn, the existence of complementarities between the preceding norms and institutions may explain the diverging path of prosperity of Europe and China (Greif and Tabellini (2017)).

Cultural evolution theories:

The first formal theoretical contributions to the modeling of the evolution of cultural preferences are due to Cavalli-Sforza and Feldman (1985) and Boyd and Richerson (1985). Using an elegant a-sexual reproduction evolutionary framework, the former hypothesized that cultural traits are both inherited - i.e. *vertically transmitted* - and transmitted by the society at large - i.e. *obliquely transmitted*. The key selective mechanism in a society composed of various cultural groups is then strength of vertical transmission, which is considered as exogenous by the former, while the latter allow for frequency dependence.

In a series of article, Alberto Bisin and Thierry Verdier (BV hereafter) introduced parental socialization choice in the theory of Cavalli-Sforza and Feldman (1985). Going beyond the biological basis of the preceding works, BV considered that the strength of vertical socialization relative to oblique socialization depends on a decision of the parent, and more specifically on a costly effort. This seemingly simple extension became a cornerstone in the study of the co-evolution of culture and societal outcomes, since it makes cultural evolution depend on a decision that can potentially be affected by institutions, cultural norms, political shocks or parents' expectations, among other key determinants of norms.

The micro-foundations of the dynamics of culture introduced by BV in the economic literature are presented in two early works of Alberto Bisin and Thierry Verdier, that we will briefly describe next. In Bisin and Verdier (2001) first, the parents' investments in increasing the strength of vertical transmission are only affected by the cultural composition of the population. More specifically, a parent will relax his own effort in inter-generational rearing when more agents share his cultural values. Of particular interest, this *cultural substitution* effect implies that parents belonging to smaller cultural groups invest more in transmitting their cultural norms vertically. This is precisely why cultural heterogeneity should be preserved

in the long-run in the dynamic model of Bisin and Verdier (2001), by contrast with Cavalli-Sforza and Feldman (1985). Bisin and Verdier (2000b) provide a thorough analysis of the cultural substitution phenomenon, and of its linkage with marital segregation decisions. As a second example, in Bisin and Verdier (2000a), the authors demonstrate that their model of cultural evolution can be particularly useful for the study of the interaction between politics and culture. Indeed, the efforts invested by the parents in vertical transmission depend on their expectation of future policies. This makes the dynamics of cultural values endogenous to the expected political equilibrium. Reciprocally, the parents vote according to their cultural values, which straightforwardly makes public policies endogenous to the cultural composition of the population.

An alternative approach to the issue of the evolution of preferences has been pioneered by Güth and Yaari (1992). In their *indirect evolutionary* approach, preferences on which rational choices depend are treated as endogenous to an evolutionary process. Yet in this process, choices are not motivated by an evolutionary success, but evolutionary success depends on the choice that are made. The indirect evolutionary approach allowed economists to approach issues such as the evolutionary stability of preferences in games with complete [2] and incomplete information [3] It also permitted to study the effect of assortative matching in repeated games on the stability of various types of preferences. [4] Wu (2016) and Besley and Persson (2017) have recently relied on the indirect evolutionary approach to study the co-dynamics of political institutions and cultural values.

The main difference between the indirect evolutionary approach and the model of BV is that the latter micro-founds the evolution of preferences in a theory of inter-generational cultural transmission, while the former does not presuppose a particular mechanism of cultural change. Indeed, in the model of BV, micro-level parental decisions of cultural transmission are affected by the macro-level socioeconomic parameters and the prevailing (and/or expected) institutions.

2. See for instance Robson (1990), Güth and Yaari (1992), Bester and Güth (1998), Alger (2010) and Alger and Weibull (2010, 2013).

3. See for instance Ok and Vega-Redondo (2001), Dekel, Ely, and Yilankaya (2007) and Alger and Weibull (2013).

4. On the stability of moral preferences, see Alger and Weibull (2016). More broadly, contributions on the subject include Alger (2010) and Alger and Weibull (2010, 2012, 2013).

This, in turn, is central in explaining the joint evolution of culture and institutions. [5]

The formation of identity:

We reach at this point the fascinating issue of the formation of preferences. This issue can not be avoided in a theory on culture and institutions, since the mechanism that determines agents' preferences also conditions the macro-micro linkage between individual decision on the one hand and institutions on the other hand. In other words, the mechanism of preference formation conditions the joint dynamics of culture and institutions. Intergenerational transmission is the technology of formation of preferences in the BV theory. Yet the micro-foundations of cultural norms remains poorly understood so far. The next paragraphs provides a brief overview of the literature on the subject.

First, there is a literature that studies evolution as a mindless process of mutation and selection akin to that studied in biology. In this literature, it is customary to consider a principal-agent framework where the principal represents the process of natural selection and the agent an individual carrying a set of genes. The principal equips the agent with some utility function so as to maximize his fitness. Evolution "hard-wires" utility functions (or preferences) that provides the goal for optimal behaviors from an evolutionary perspective, along with a learning process that would help to pursue that goal (Robson (2001); Becker and Rayo (2007b); Robson and Samuelson (2011)). As a simple illustration, such a mechanism could explain why we attach utility to activities such as eating. [6] Becker and Rayo (2007b,a) suggest that utility functions that measure individual's success in relative terms and which reference points change according to the agent's performance are fitness maximizing. The preceding theory could then explain the importance of habits and peer comparisons in human societies, as well as series of result in the subjective well-being empirical literature demonstrating that

5. In that respect, it is interesting to observe that the recent approaches on the joint evolution of culture and institutions of Wu (2016) and Besley and Persson (2017) seemingly established under the indirect evolutionary approach also micro-found the interaction between culture and institutions in a theory of intergenerational cultural transmission.

6. Important contributions to this literature include Robson (2001), and Samuelson (2001), Samuelson (2004), Samuelson and Swinkels (2006) and Robalino and Robson (2012). See as well the review of Robson and Samuelson (2011) and the references therein.

level of happiness continuously reverts to its long-term mean.

A second line of work suggests that identities are adopted because they are self-serving. Identity can grant self-esteem because it is an informative signal of congruence with group values (Bénabou and Tirole (2006b, 2012); Akerlof (2016)). Respecting the costly requirements of a religion signals that one has "what it takes" to be religious for instance. In the spirit of Bénabou and Tirole (2011), a related view consists in arguing that people have a better record of their past behaviors than their past motivations, so the questions "who we are" and "what we do" tend to be assimilated. Relatedly identities can be viewed as commitment devices that make people adopt most often the behaviors that correspond to their cultural identity (see for instance the theory on veiling of Carvalho (2013)).

Finally, and in accordance with the central mechanism of the BV theory of cultural evolution, only two works to my knowledge in the economic literature study the formation of preferences through inter-generational rearing, Lizzeri and Siniscalchi (2008) and Doepke and Zilibotti (2017). For the former, parenting strategies arise from limits on communication and interpretation. In their model, parents intervene to protect children from the consequences of ill-informed choices. This, in turn, reduces children's ability to learn from experience. For Lizzeri and Siniscalchi (2008) then, alternative parenting strategies arise depending on parents' willingness to shelter their children from the negative consequences of bad choices. In Doepke and Zilibotti (2017), the authors aim to explain variation in parenting styles over time and across countries. For that purpose, they formalize parenting styles by following the classification established in the developmental psychology by Diana Baumrind (1967). They consider that authoritative parenting styles affect preferences while authoritarian parenting methods constrain children's actions. Their theory is built on the idea that inter-generational disagreements that have a bearing on human capital investment and economic success are central in understanding observed variations in parenting styles. Parents can then force behaviors on their children through authoritarian strategies when it is supposed to significantly increase children's economic success in the future. As an illustration, in equalitarian societies, the returns from pushing children in one specific direction may be relatively low. According

to the author, this could explain why a decline in authoritarianism is observe in the 1960s and the early 1970s, a period characterized by historically low economic inequalities.

The joint evolution of culture and institutions:

The initial theories of Alberto Bisin and Thierry Verdier allowed to study a plethora of issues (see the review of Bisin and Verdier (2011) for instance). Of particular interest to the topics of this dissertation, I will now describe the recent contribution of Bisin and Verdier (2015) on the joint evolution of culture and institutions. In the paper, the authors seek to understand the interactions between culture and institutions, and more specifically under which conditions the joint evolution of culture and institutions lead to desirable socio-economic equilibria. The authors conceptualize institutions as mechanisms through which social choices are delineated and implemented. The central novelty of their approach is to consider that institutional change aims at internalizing the lack of commitment and the externalities which plague social choice problems. Institutional change is then by nature incremental rather than discontinuous in Bisin and Verdier (2015). Their theory is general enough to accommodate for previous attempt to study the dynamics of institutions (e.g. the pathbreaking works of Daron Acemoglu and James Robinson), while it opens new perspectives on development trajectories of various societies.

This thesis: from the formation of preferences to the determination of political institutions

While scholars have identified important determinants of norms and institutions in recent years, there is still a deep remaining conundrum, as argued recently by Kranton (2016, p. 407): "why do divisions and norms resonate for human beings?" In the jargon of economists, what are the micro-foundations of identity? This interrogation motivated the first essay of this dissertation, which looks upstream of the relationship between culture and institutions

and studies the formation of cultural identity through intergenerational interactions with caregivers.

Few works have studied the formation of cultural identity, as reviewed earlier and even fewer have considered the importance of intergenerational rearing in the process. The influence of the parents on the formation of cultural identity seems nevertheless paramount. In a systematic study of the effect of the variability of ecological conditions across generations, Giuliano and Nunn (2016) provide empirical validation of a class of models that hypothesis that it optimal to rely on cultural norms when making decisions and that parents are able to influence the preferences of their children. More broadly, Bisin and Verdier (2011) provide empirical references on the persistence of cultural values across generations. Another important motivation for studying the interaction between the formation of identity and intergenerational rearing relates to the very nature of human development. There is compelling evidence that the early years of life are critical and sensitive periods in the development of the brain (Cunha and Heckman (2007)). [7] Furthermore, identity imprignates our brain - as evidence in neuroscience suggests. [8] and it is precisely during the critical period for the development of the brain that parents are the most important interacting peers of young infants. A satisfactory theory of the formation of identity should account for the peculiarities of an intergenerational mode of transmission.

The first chapter presents a theory on the formation of identity that begins with a simple model of child-caregiver interactions. In the model, the parents chooses the magnitude of a pecuniary reward and the margin of error they authorize in their children's behavior. Children decide the level of effort they invest in rearing interactions. This modeling choice allows to link parental rearing strategies to the taxonomy of parenting styles established in the developmental psychology literature (Baumrind (1978)). Furthermore, as a child learns to recognize the states of the world that are associated with a given cultural norm, he develops a

7. There is considerable evidence for instance that the later the remediation to a disadvantaged child, the less effective it is. See for instance Cunha and Heckman (2007, 2008); Cunha, Elo, and Culhane (2013); Heckman and Corbin (2016).

8. See for instance

capital of appreciation for behaving according to that norm.[9] This creates complementarities in the formation of identity and the accumulation of cognitive skills, in accordance with series of evidence in the literature on cognitive development (e.g. Cunha, Elo, and Culhane (2013); Heckman and Corbin (2016)).

This theoretical framework allows to make interesting predictions on the correlates of parenting styles and of the accumulation of a cultural capital during the early years of life. We can also approach fascinating issues that traditionally relied outside the reach of economists such as child maltreatment and children television viewing. A first interesting prediction of the this essay is that the time preferences of the parents are critical in explaining their ability to set efficient rearing strategies. Indeed, a parent with higher time preferences will be able to anticipate the existence of dynamic complementarities in her child's accumulation of cognitive/cultural capital. She may then be particularly involved in early rearing interactions so as to put her child on the best accumulation path. As a simple illustration of this prediction, Rowe (2008) argues that gaps in early home language environments exist because poor, uneducated mothers do not know about the role they play in determining the language and cognitive development of their children.

Relative to the parenting styles induced by poverty, I show that it goes from higher degrees of authoritarianism to child neglect and can even lead to child maltreatment. Poverty induces authoritarian parenting styles because poorer parents substitute their lack of investments in child rearing by reducing the margin of error they authorize in their children's behaviors. Simply put, they substitute the benefits of the carrot for the fear of the stick. But the compensation is only partial and authoritarianism leads to lower investments of the child in the rearing interactions in equilibrium. At the extreme, child maltreatment arises when the parents fail to recognize the long-term consequences of their influence on their children's development. They do not foresee the influence that their behavior has on their offsprings' production of toxic stress and healthy brain development. I find that maltreatment is likely to be persistent when it arises. This result accords for instance with the evidence of Proctor, Aarons, Dubowitz,

9. This assumption is motivated by the works of Gary S. Becker and co-authors on the acquisition of cultural capital.

English, Lewis, Thompson, Hussey, Litrownik, and Roesch (2012) that maltreated children between ages 4 and 12 are highly likely to be abused or neglected in the future, absent intervention.

Coming back to the common theme of the essays of this dissertation, culture and institutions, it has been recently argued by Bisin and Verdier (2015) that knowing the *origin* of either culture or institution loses much of its interest since it is the joint evolution of culture and institutions that determines economic growth and prosperity. In their own words, "the focus is moved from the cause (both culture and/or institution can have causal effects) to the process that determine the interaction" Bisin and Verdier (2015, p. 3). The second chapter of this dissertation studies the joint evolution of a religious identity and economic and political institutions.

One important motivation of this chapter is the observation that religious cultures are not neutral with respect to economic activities. Both Muslim and Christian faiths have imposed bans on usury activities. Scientific and innovative activities encountered restrictions as well and technology adoptions were repeatedly delayed throughout history. In the second essay, I seek to understand why at some point in their evolution religions can become hostile to some economic activities. To this end, I develop a model of cultural evolution in a productive economy with endogenous innovation.

In the basic version of the model, a religious leader controls the doctrine and decides whether to discriminate some economic activities. The agents allocate their labor in one of two productive sectors and there exist monopoly producers of sector-specific technologies. One sector is potentially subject to prohibition because it aggregates the occupations that are complementary to scientific discoveries for example. In the former case, the monopoly producers of technologies in the discriminated sector are scientists and innovators. The sector that is never subject to prohibition can be thought of as encompassing traditional occupations within the religious communities.

The model of production with endogenous innovation is embedded in a cultural evolution framework. I assume that instead of responding to the "eroding" effect of scientific innovation

on religious beliefs in a given period (Bénabou, Ticchi, and Vindigni (2015)), the religious leader internalizes the entire dynamics of religious norms. [10] Prohibition aims at altering the dynamics of religious preferences in a way that maximizes the rents of some religious leader. The central intuition of this chapter is that prohibition creates a cultural segmentation of the labor market. The Christian or the Muslim prohibition against usury should pave the way to some related occupations to religious minorities. Furthermore, a culturally segmented labor market affects the dynamics of cultural norms because it differentiates cultural groups. In turn, such an "economic" distinction between cultural groups adds a salient dimension to parents' existing incentives to transmit their cultural norms intergenerationally. This is why prohibition can substantially increase the diffusion of religious preferences in the population, and can therefore substantially benefit religious leaders.

In this chapter, I show that there is a timing in the instigation of prohibition. It will eventually arise if the size of the religious group is high enough, so that discriminations affect the labor allocation on the first hand, but not overwhelmingly high, so that the induced sorting effects favor the spread of the religious trait on the other hand. Prohibition is (ironically) a "doctrinal innovation" that allows strong religious preferences to take root in the population. Relative to this result, In the Islamic world, Chaney (2016) shows that a significant drop in books written on scientific topics becomes statistically significant in the twelfth century. [11] Relatedly, regarding prohibition against usury activities in Europe, the Catholic Church was primarily forbidding them to clerics in the 500-1050 period (Reed and Bekar (2003)). Usury became a dominant concern only during the 1175-1350 period, with a peak at the Council of Lyon in 1274 and at the Council of Vienne in 1312 (Reed and Bekar (2003)). The enforcement of the prohibition was relaxed only by 1830, when "the Sacred Penitentiary issued instructions to confessors not to disturb penitents who lend money at the legal rate of interest." (De Roover (1974, p. 321) quoted by Reed and Bekar (2003)).

10. This modeling choice follows the recent research of Verdier and Zenou (2016) on the interaction between centralized and decentralized evolutionary process.

11. Chaney (2016) proxies the scientific production in the Muslim pre-industrial world with Harvard's library holdings.

The regulation of political competition is a major characteristic of political institutions. Yet the process of selection of political rulers is more than often modeled by political economists as a two-party contest for power. This simplification overshadows the study of the socio-economic determinants of political competition, and therefore constrains to a large extent the analysis of the selection of political institutions. As a simple illustration, the study of the interaction between the cultural composition of the population and the feasible political institutions remains a largely open research area. The third essay of this dissertation, which is a common work with Pr. Thierry Verdier, aims at filling this gap.

The purpose of the last essay is to provide a simple analytical framework on multi-candidate elections that is tractable enough to allow for a systematic analysis of the endogenous structure of political competition in ways similar to the Industrial Organization models used to analyse endogenous market structures under economic competition. The versatility of our approach is demonstrated through several applications on classical topics in political economy: special interest politics, coalition formation in the legislature, and franchise extension in the democracy.

The model is a standard probabilistic voting theory (e.g. Coughlin (1992) and Persson and Tabellini (2002)). We consider however that the noise in random voting decisions is distributed according to some Fréchet (or extreme type II) distributions. We do so because the maximum of a finite sequence of random variables distributed according to Fréchet distributions is a contest function. This idea has already been exploited in the context of trade between multiple countries by Eaton and Kortum (2002). It is particularly relevant in the context of political competitions as well, since individuals cast their vote for their most preferred candidate out of a finite list of challengers. Our approach implies a simple formalism for the determination of electoral equilibria and for dealing with the issue of the endogenous entry of candidates in election.

We also outline an axiomatic approach to probabilistic voting models that provides a micro-founded rationale for the use of Fréchet distributions in voting theories. The analysis is inspired of the seminal approach of McFadden (1974) of individual choice decisions. We assume that in a probabilistic voting model, voting behaviors follow three Axioms. Those

axioms are respectively the Independence of Irrelevant Alternatives, the positivity and the Irrelevant of Alternative Set Axioms (McFadden (1974)). As an illustration of our result, assume that the voters use a random voting model. We seek to understand what kind of distribution for the noise in voting decisions is consistent with the three preceding axioms and find that only the Fréchet distributions satisfy them.

Among the key contributions of the paper, we establish that both proportional and run-off systems should lead to higher party fragmentation than a plurality system. Furthermore, we are able to show that allowing high campaign investments leads to less fragmented polities and more rent extraction from well funded political parties.

Chapter 1

Parental Rearing Practices, Cultural Transmission and Cognitive Development[1]

1. I am grateful to Thierry Verdier, this paper owes him much. Thanks also to Gani Aldashev, Emeline Bézin, Alberto Bisin, Itzhak Gilboa, Rachel Kranton, Sultan Mehmood and Ragnar Torvik. Financial support from the European Research Council under the project Tectacom 324004 is gratefully acknowledged. All mistakes are my own

Abstract

This paper presents a theory of child development and parental rearing practices. In the model, a benevolent parent seeks to transmit behavioral norms to her child by sending him signals that are imperfectly observed. The child can however increase the quality of the signals he receives by investing in acquiring cognitive skills. We establish that neither authoritarian nor permissive parenting styles are conducive to the accumulation of cognitive skills. Furthermore, since rearing interactions aim at transmitting behavioral norms, the child develops a capital of appreciation for particular cognitive skills. This cultural perspective to the issue of cognitive development provides an interpretation grid for various results established in the empirical literature on child development. Our approach also permits to identify the parental characteristics that are particularly conducive to child maltreatment. We predict that child maltreatment should be expected to emerge early in the rearing period and to be persistent, hence the need for early and repeated interventions for maltreated children. We also suggest that the use of television and video materials for rearing purposes can widen the existing disparities in the early acquisition of cognitive skills.

JEL codes: D10, D91, Z10.

Keywords: Parenting styles, Cognitive Development, Intergenerational transmission, Cultural Transmission.

If a child asked me whether the earth was already there before my birth, I should answer him that the earth did not begin only with my birth, but that it existed long, long before. And I should have the feeling of saying something funny. Rather as if the child had asked if such and such a mountain were higher than a tall house that it had seen. In answering the question I should have to be imparting a picture of the world to the person who asked it.
If I do answer the question with certainty, what gives me this certainty?

— Ludwig Wittengstein, *On Certainty (par. 233)*

1.1 Introduction

In recent years, an increasing amount of attention has been devoted to the formation of cognitive skills during the early years of life. Indeed, inequalities in cognitive skills emerge early and have long lasting effects on education, earnings, employment, crime, marriage and participation in healthy behaviors among other key socio-economic outcomes. [2] At the center of this research endeavor lies the study of the effects of the parental environment and parents' investments in child rearing at different stages of childhood. [3] Yet no work in the literature to our knowledge formally studies the peculiarities of rearing practices and their effects on child development. This paper aims at filling this gap.

To this end, we build a theory of child development that accounts for rearing strategies. The theory begins with a simple model of child-caregiver interactions, which translates the idea that children are presented with stimuli and asked to accomplish goals formulated for them by upbringers. Upon observing the state of the world, a single parent sends a signal to her child so that he will be able to behave accordingly. The child imperfectly observes the signal sent by her parent. Yet he can invest resources in order to get a more precise interpretation of what he observes. This investment of the child is our proxy for cognitive development. In order to encourage her child to invest resources in developing his cognition, the parent has two levers. First, she provides a caring environment and expresses her love and affection. Second, she chooses when to show her affection, i.e. she decides the margin of error tolerated in the child's behavior. For instance, a parent that is permissive will always show her love and care to the child, even if the child's behavior diverges significantly from what is initially expected of him. By contrast, a parent that is more authoritarian will condition significantly more her expressed level of affection on the behavior of the child. This simple dual decision of parents allows to relate our theory of child rearing interactions to the three prototypical

2. See for instance Campbell and Ramey (1994); Heckman, Moon, Pinto, Savelyev, and Yavitz (2010); Walker, Wachs, Grantham-McGregor, Black, Nelson, Huffman, Baker-Henningham, Chang, Hamadani, Lozoff, Gardner, Powell, Rahman, and Richter (2011); Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014); Campbell, Conti, Heckman, Moon, Pinto, Pungello, and Pan (2014) and Conti, Heckman, and Pinto (2016).

3. See the seminal work Cunha and Heckman (2008) for example and the review of Attanasio (2015).

parenting styles identified in the seminal work of Baumrind (1966). We have described above the permissive and authoritarian parenting styles respectively. The last category, labeled authoritative, relates to a balance of authoritarianism and permissiveness in child rearing.

One important novelty of our approach is to embed our theory of child development through rearing interactions in a broader framework of intergenerational cultural transmission. Indeed, we consider that the knowledge that is transmitted from the parent to the offspring in rearing interactions is cultural. That is to say, each behavior that the child learns is thought as being associated to a *cultural perception* of the parent. Take the example of the norm patience. It relates to the delaying of present activities, say consumption, given that one foresees the long-run benefits of the delay. Foreseeing the future stream of benefits of the delay is the *cultural perception* associated with the norm patience. It is no mystery that in plenty of situations, a parent has to repeatedly teach her child to delay his present consumption. More broadly, many aspects of intergenerational learning and child development are cultural. Language is a prototypical example. The acquisition of language skills constitutes a prominent part of early interactions between toddlers and caregivers and corresponds to the way previous generations of human beings in a given location or social group have found it useful to categorize and structure the world (Tomasello (2009, Chapter 3)). Language skills are also standard measures of cognitive development.

We exploit the close linkage between child development and cultural transmission by assuming that as the child learns to behave according to a given cultural norm, he develops a capital of appreciation for behaving according to that norm. Dynamic complementarities then arise in the acquisition of cognitive skills in this model because child rearing interactions allow for the emergence of a cultural capital. Furthermore, this framework allows to make interesting predictions on the correlates of parenting styles and of child develeopment. We outline some of the most important results of our study in the next paragraphs.

First, we show that the time preferences of the parents are critical in explaining their ability to set efficient rearing strategies. Indeed, a parent with higher time preferences will be able to anticipate the existence of dynamic complementarities in her child's cognitive

development. Thus, she may be particularly involved in early rearing interactions so as to put her child on the best accumulation path. As a simple illustration, Rowe (2008) argues that the gaps in early home language environments exist because poor, uneducated mothers do not know about the role they play in determining the language and cognitive development of their children. In our view, this may be explained by the lower time horizon of uneducated women since it is precisely what makes them underestimate their influence in their children accumulation of skills.

Furthermore, an efficient parenting style is neither permissive nor authoritarian but authoritative. Excessive permissiveness trumps out the child's investments in recognition because the latter always get rewarded by the love of his parent, even if his behavior deviates significantly from what is expected of him. Alternatively, the high conditionality of love and care associated to authoritarian parenting styles creates a weak incentive for the child's investments in rearing interactions. This simple result accords with series of evidence showing that either authoritarian or permissive parenting styles lead to poorer cognitive outcomes (see Chan and Koo (2011) and Piotrowski, Lapierre, and Linebarger (2013) for example).

We establish that poverty has negative and long lasting effects on the child's investments in recognition while it favors unhealthy parenting styles. The negative linkage between income and children cognitive outcomes is well documented. As a simple illustration, Rubio-Codina, Attanasio, Meghir, Varela, and Grantham-McGregor (2015) find that the significant differences that emerge in cognitive and language development among children of different socioeconomic backgrounds at age 12 months grow substantially over time (see as well the insightful review of Attanasio (2015)). Relative to the parenting styles induced by poverty, it goes from higher degrees of authoritarianism to child neglect and can even lead to child maltreatment, as we suggest in the first extension of this paper. Poverty induces authoritarian parenting styles because poorer parents spend a lower fraction of their time expressing love and care and consequently condition relatively their love on the behavior of their offspring. Simply put, poorer parents substitute the benefits of the carrot for the fear of the stick. But the compensation is only partial and authoritarianism leads to lower investments of the child in

the rearing interactions in equilibrium. Additionally, the accumulation of low investments over time creates gaps in cognitive skills between the children belonging to different income groups.

Going further on the issue of unhealthy rearing practices, our first extension introduces the use of maltreatment in child rearing as well as the production of toxic stress during childhood. Child maltreatment is thought as a rearing strategy that is costless for the parent. Child maltreatment then contrasts with the previous rearing practices, since providing a caring environment for the child takes time and necessitates monetary investments. We also posit that maltreatment depreciates the cognitive capital that is accumulated by the child. This modeling choice is motivated by the evidence showing that child maltreatment and child neglect induce the production of toxic stress, which damages brain architecture by leading to the underdevelopment of certain parts of the brain that are necessary for emotional control, memory, learning and problem-solving (National Scientific Council on the Developing Child (2005a), Shonkoff, Garner, Siegel, Dobbins, Earls, Garner, McGuinn, Pascoe, and Wood (2011)). We show that child maltreatment arises when the parents fail to recognize the long-term consequences of their influence on their children's development, when their marginal utility from private consumption is high, or when they are weakly altruistic toward their children. Furthermore, besides having negative effects on the accumulation of cognitive skills, we show that when maltreatment appears early in childhood, it will persist throughout the rearing period. In the United States, the proportion of children who experience a report to Child Protective Services and that are re-reported is approximately equal to 25% according to the U.S. Department of Health and Human Services Administration on Children Youth and Families (2010). Our result also accords with the evidence of Proctor, Aarons, Dubowitz, English, Lewis, Thompson, Hussey, Litrownik, and Roesch (2012) that maltreated children between ages 4 and 12 are highly likely to be abused or neglected in the future, absent intervention. Furthermore, young child age has emerged as a key risk factor. Compared to other age groups, children under the age of five have the highest rates of maltreatment reports (U.S. Department of Health and Human Services Administration on Children Youth and

Families (2010)) and are at the highest risk for re-reports (Bae, Solomon, and Gelles (2009) and Fluke (2008)).

Finally, we are able to use the framework set in this paper to study how television and video materials affect child rearing strategies and cognitive outcomes. We believe this issue to be particularly important because there has been an explosion of videos and television viewing in the lives of infants and toddlers in recent years. The American Academy of Pediatrics (AAP) recommends no screen time for children younger than 2 years (The American Academy of Pediatrics (2001)), although this recommendation is largely ignored (Certain and Kahn (2002); Zimmerman, Christakis, and Meltzoff (2007a)). Our key takeaway in that respect is that television viewing, if not compensated by a higher involvement of the parents in child rearing, will lead to decreasing investments of children in the acquisition of cognitive skills. Television viewing can then have deleterious consequences on children cognitive development when it does not increase the parents' own investments in child rearing. One important consequence of this is that poorer parents will let their children watch too much television when altruistic as they will not be able to invest more resources in child rearing. Televisions and video materials should then increase disparities in the acquisition of cognitive skills between income groups. Although the literature on the subject is still in its infancy, some studies go in this direction and find that television viewing correlate with socio-economic status, and that the effect of television viewing is significant and negative for the infants and toddlers belonging to the lower socio-economic strata (Zimmerman and Christakis (2005); Zimmerman, Christakis, and Meltzoff (2007b); Christakis, Zimmerman, DiGiuseppe, and McCarty (2004)).

1.1.1 Related literature

This paper relates to the rich literature on child development. Our main contribution to this literature is to account systematically for the linkage between parental rearing practices and child development. This generates interesting predictions on how parental characteristics affect child development. Furthermore, as summarized above, this allows to study interesting issues that traditionally lied outside the reach of economics such as child maltreatment and children

television viewing. Furthermore, our cultural approach to child development is consistent with recent series of evidence that the long-term impact of programs such as the Abeccedarian project, the Perry Preschool program, Head Start or Project STAR is due to increases in skills that are not measured by grades or IQ tests (Algan, Beasley, Vitaro, and Tremblay (2014)). Rather, those programs typically affect social behaviors by increasing the participation in healthy behaviors, enhancing adult outcomes including education, employment, earnings, marriage, and reducing the participation in crime. Heckman and Kautz (2013) have suggested that what they label character skills - a concept derived from the personality psychology literature - are important ingredients of successful lives, affect positively the acquisition of cognitive skills - and reciprocally - and can be acquired during childhood. The concept of character skills is closely related to that of personality traits in the personality psychology literature. [4] Character skills include perseverance, self-control, trust, attentiveness, self-esteem and self-efficacy, resilience to adversity, openness to experience, empathy, humility, tolerance of diverse opinions, and the ability to engage productively in society (Heckman and Kautz (2013)). They are shown by various studies to have particularly positive effects in the success of live, work or education. [5] Furthermore, several studies suggest that interventions during the preschool years or in kindergarten improve character skills in a lasting way and that they explain a substantial share of the long run effects of early intervention programs (see Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) for example).

Second, this paper relates to the emerging economic literature on child rearing and parenting styles. Our main contribution to this literature is to link the study of intergenerational rearing to that of child development. Our canonical model of child rearing is inspired of the contributions of Weinberg (2001) and Lizzeri and Siniscalchi (2008) in that we also consider a principal-agent framework. However, we consider a two-dimensional decision for the principal (the parent). Indeed, she controls both the level of reward when the child behaves well and the margin of error. By doing so, we can closely relate the decision of the parent to

4. See the reviews of Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007), Almlund, Duckworth, Heckman, and Kautz (2011) and Heckman and Kautz (2012).

5. See for instance Heckman and Kautz (2012, 2013) and the references therein.

the parenting styles identified by Baumrind (1967). Furthermore, we introduce a dynamic model of child rearing that accounts for the accumulation of a cultural/cognitive capital through repeated rearing interactions. This part of the model closely relates to the seminal approach of Becker and Murphy (1988) to the formation of an appreciation capital. The most closely related work in the literature on child rearing is Doepke and Zilibotti (2017). The authors formalize as well parenting styles by relying on the classification of Baumrind (1967). Their approach differs from our's in that they consider that authoritative parenting styles affect preferences while authoritarian parenting methods constrain children's actions and do not affect preferences. Our approach to parenting styles does not assume such a distinction but rather starts from the premise that the objective of early intergenerational rearing is to transmit preferences. Furthermore, we consider a continuum of parenting styles from permissive to authoritarian methods, with authoritativeness being defined as a balance of permissiveness and authoritarianism. Additionally, while Doepke and Zilibotti (2017) focus on the macro-determinants of parenting styles such as income inequalities, we are interested on the micro-determinants of rearing practices and of the acquisition of cognitive capital. Finally, our dynamic theory goes beyond the existing works in this literature because it allows to study the evolution of rearing practices and the formation of cognitive skills.

We finally relate to the literature on the formation of preferences. [6] Our main contribution to this literature is to propose a definition of cultural norms that uses the concept of Aumann structure (see (Fagin, Halpern, Moses, and Vardi (2004, p. 37)) and Aumann (1959)). This definition of culture is the cornerstone of the study of the formation of preferences through repeated child rearing interactions.

The next section presents a canonical model of child rearing interactions. The dynamic model that accounts for child rearing and the accumulation of a cognitive capital is presented in the third section. The fourth section presents two extensions of the theory and the fifth section concludes. The proofs of the propositions are relegated to the Appendix.

6. See for instance Akerlof and Kranton (2000), Robson (2001), Bisin and Verdier (2001), Akerlof and Kranton (2010), Rayo and Robson (2013) and the reviews of Bisin and Verdier (2011), Robson and Samuelson (2011) and Kranton (2016).

1.2 A static model of child rearing

In this section, we present a canonical model of cognitive development through intergenerational rearing interactions. We define cognition by an Aumann structure (Fagin, Halpern, Moses, and Vardi (2004, p. 37)) enriched by reasoning abilities and action rules contingent on this structure. As defined in Fagin, Halpern, Moses, and Vardi (2004, p. 37), an Aumann structure is a tuple $(\mathcal{S}, \mathcal{P}_1, \ldots, \mathcal{P}_m)$ where $\mathcal{S}$ is the set of states of the world and $\mathcal{P}_i$ is a partition of $\mathcal{S}$ for some agent i. The cognition of an agent corresponds to two things. First, a representation of the world by a finite set of possible states of the world. Any state of the world is empirical in the sense that any w_k^i can be defined as "the state of the world where action $a(w_k^i)$ must be performed". We assume the set of possible actions $\mathcal{A}$ to be convex and bounded. Second, the cognition of an agent implies reasoning. Empirical frequencies for instance can be built upon the knowledge structure and the varied experiences so as to make Bayesian reasoning feasible.

The preceding formalization encompasses the definition of cognition of the American Psychological Association Dictionary, "all forms of knowing and awareness such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving." Awareness, perception, conception and memory subsume to a subjective structure of knowledge while reasoning, judging and problem solving relate to the ability to link the subjective states and act contingently. As a simple illustration, a state in the knowledge structure can be "it is raining". Two agents can differ in their evaluation of the state, as well as in their action upon having the same knowledge. In that example, rain can be associated with a state "the Gods are angry" for one (which could also be associated with other meteorological events) and alternatively with "this is a meteorological event called rain" for another for instance. Those two events will most certainly trigger different behaviors. Knowledge structures are subjective and give the prism of agents' perception.

1.2.1 The child-rearing interaction

Consider now a parent engaging in early inter-generational rearing. I will use the pronouns she for the parent and he for the child. For the parent, there are n_τ states of the world that are worth being transmitted. We can think of those states as corresponding to the transmission of language and standard norms of behavior for example. Thus, the set of states of the world that are worth being transmitted is a subset of the knowledge structure of the parent, $\Omega^\tau = \{\omega_1, \ldots, \omega_{n_\tau}\} \subseteq \Omega$ with Ω the knowledge structure of the parent. Instead of making specific assumptions on the structure of reasoning, I will simply assume that for the parent, for each state ω_k^τ that is worth being transmitted, there exists only one best possible action for the child $a(\omega_k^\tau)$.

Consistently with our previous discussion, through rearing, the child has the ability to distinguish experiences and build subjective concepts. Furthermore, the child knows that his caregiver as well distinguishes subjective states of his own, given that he apprehends her as an intentional being. Thus, we can formalize a rearing interaction between the caregiver and the child as a principal agent model where the parent seeks to mold the cognition of her child.

Before turning to the transmission of Ω^τ, we highlight the underpinnings of the transmission of a single state $\omega_k \in \Omega^\tau$. When the state of the world is w_k, the parent believes that the correct action to perform a^* is distributed according to a normal distribution centered on $a(w_k)$ with precision s_p. The ability to know that when the state is w_k the correct action is normally distributed with mean $a(w_k)$ could be the result of the parent's own building of state recognition during her childhood for example. Similarly, we assume that a priori, the child believes that the correct way to act in state w_k is normally distributed with mean $a(w_k)$ and precision $f(e)$. Contrary to the parent however, the child does not know $a(w_k)$. He has a prior on $a(w_k)$ that is normally distributed with mean a_0 and precision s_0. The main novelty of our framework is to consider $f(e)$, the precision of the child's own beliefs on the correct action, to be endogenous to his recognition effort e invested during the rearing interaction. In other words, the precision of the child's recognition of the best possible action - and thus of the state of the world - depends on his own investments during the rearing interaction.

Thus, the rearing strategy of the parent will aim at setting an optimal reward scheme so as to spur the child's investment during the interaction. This simple model of rearing interactions translates the idea that children are presented with stimuli and asked to accomplish goals formulated for them by upbringers.

In the static model of this section, the child's investment in state recognition is our only proxy for cognitive development. Although we elaborate more on cognitive outcomes in the dynamic model of the next section, observe that e relates to various standard dimensions of cognitive abilities. Indeed, if we think of a child rearing interaction aiming at transmitting language skills for example, then a child that invests a higher recognition effort will be better able to pronounce the words he hears, to recognize them in sentences and to understand their meaning. He may as well develop higher memory and listening skills. Finally, a child that invests in recognition will have a higher psychosocial development, since he will be better able to act appropriately to the social standard he learnt from caregivers. Thus, we should expect a child that invests in recognition to perform better on standard cognitive tests.

The timing of the game is the following. The parent knows that the correct action to perform for her child that we denote a^* is normally distributed with mean $a(w_k)$ and precision s_p, so she believes that w_k is the most likely state of the world. First, she commits to a the level of love and care and decides as well which actions will deserve her love and care. Anticipating the parental rearing strategy, the child chooses his effort in improving the precision of his signal e. The rearing interaction then occurs. We model the rearing interaction as if the parent was transmitting a signal to the child a^{obs} on the nature of the state of the world. The parent believes the state of the world to be w_k with the highest probability. She then tries to send $a(w_k)$ during the rearing interaction but the child incorrectly observes his parent's signal. Instead of receiving $a(w_k)$, the child observes a^{obs}, which is drawn from a normal distribution centered on $a(w_k)$, with a precision $f(e)$ that is endogenous to his effort e. Finally, the child chooses an action to perform and the parent implements her rearing strategy.

We assume that the child is endowed with a recognition technology $f(.)$ that is increasing, concave, not state-specific and that depends on his recognition effort e, with $f(0) = 0$. The

higher e, the more precise is the child understanding of the state of the world. Observe that since the signal is transmitted by the parent - who has her own beliefs on the state of the world - there is an auto-corellation of knowledge structures across generations in this model. Furthermore, by contrast with Lizzeri and Siniscalchi (2008), we assume that the parent does not deviate from the truth in the sense that the mean of the signal received by the child out of his interaction with his caregiver is $a(w_k^i)$, the best possible action in state w_k^i from the parent's perspective. We make this assumption for simplicity. Consistently with Lizzeri and Siniscalchi (2008), the parent could deviate from his most preferred action so as to protect her child from the harsh consequences of a bad choice.

Following Cyert and DeGroot (1987, p. 19) for example, it can be shown that the posterior distribution of the child on the best way to act in state w_k (i.e. on $a(w_k)$) is normally distributed with precision $s(e) = s_0 + f(e)$ and mean

$$\overline{a}(\omega_k) = \frac{a_0 s_0 + a^{obs} f(e)}{s_0 + f(e)}. \tag{1.1}$$

We assume that the parent affects the utility of her child by expressing love and care. More precisely, we posit that the parent can give a fixed utility δ to her child when caring for him. However, affection is expressed conditionally on the action of the child. When action $b(w_k)$ is performed by the child in state w_k, the parent expresses her affection if $b(w_k) \in [a(w_k) - \epsilon, a(w_k) + \epsilon]$ and does not otherwise. Thus - in the absence of other incentives and upon observing a^{obs} - the child will choose action $\overline{a}(\omega_k)$ during the rearing interaction because it maximizes his chance of being rewarded.[7] We will refer to δ as a reward in the sequel.

An arbitrarily low value of the parameter ϵ for a constant and positive δ refers to an authoritarian parenting style because then the parent's rearing strategy does not allow for much variations around her optimum $a(w_k)$. Indeed, according to Baumrind (1966, p. 890), an authoritarian parent "values obedience as a virtue and favors punitive, forceful measures

7. Observe that the parent can not predict $\overline{a}(\omega_k)$ when choosing her rearing strategy because she does not observe a^{obs} ex ante.

to curb self-will at points where the child's actions or beliefs conflict with what she thinks is right conduct." Alternatively, rearing strategies that favor significantly high values of ϵ can be characterized as permissive in the classification of the preceding author. Baumrind (1966, p. 889) writes "[a permissive parent] allows the child to regulate his own activities as much as possible, avoids the exercise of control, and does not encourage him to obey externally defined standards." Finally, the last type of parental control identified by Diana Baumrind is labeled authoritative and is a combination between permissive and authoritarian parenting styles, she writes "both autonomous self-will and disciplined conformity are valued by the authoritative parent" (Baumrind (1966, p. 891)). The preceding author argues that the Montessori method exemplifies authoritative rearing practices since teachers in Montessori schools exert authority but encourage as well children to engage in activities of their choosing. In this model, the parents that balance reward and permissiveness in their rearing strategy will be referred to as authoritative.

We first study the optimal strategy of the child and then turn to the strategy of the parent. We drop the index w_k when not necessary in the sequel. In state w_k, the utility of the child is

$$E_{a^{obs}} v_c(e) = \delta \, \mathbb{E}_{a^{obs}}[\int_{\overline{a}-\epsilon}^{\overline{a}+\epsilon} g(u,e)du] - c(e) \tag{1.2}$$

with g the density of the posterior distribution of the child,

$$g(u,e) = \sqrt{\frac{s(e)}{2\pi}} \exp(-\frac{1}{2} s(e)(u-\overline{a})^2) \tag{1.3}$$

The integral in (3.2) gives the probability of receiving the reward given the quality of the signal $s(e)$ that is positively affected by the recognition effort of the child e. The function $c(.)$ relates to the cost for the child of investing resources in improving the quality of the signal. We will assume that $c(e) \equiv e^2/2$ hereafter for simplicity. Observe that the integral in (2) is independent from $\overline{a}$ and simply rewrites

$$\int_{\overline{a}(w_k^i)-\epsilon}^{\overline{a}(w_k^i)+\epsilon} g(u, e_k^i)du = \int_{-\epsilon}^{\epsilon} \sqrt{\frac{s}{2\pi}} \exp(-\frac{1}{2} s u^2)du. \tag{1.4}$$

Thus, the child's optimization problem is independent from the signal he receives. Intuitively, the child's reward depends simply on the deviation between his action and the mean of his posterior. But his action is precisely the mean of his posterior, so his expected reward only depends on the quality of the signal he receives, that in turn depends on his own effort in the interaction e. This remark is important because it implies that the child needs not to observe the signal a^{obs} when choosing his recognition effort e, which is consistent with the timing previously given. We show the following result in the Appendix.

Proposition 1.

— *In any state w_k, and for a given rearing strategy of the parent (δ, ϵ), there exists a unique optimal recognition effort e for the child that aims at improving the quality of the signal he receives on the state of the world. The first-order condition associated with the child's optimization writes*

$$-e + \frac{\delta \epsilon f'(e)}{\sqrt{2\pi(s(e)}} \exp(-\frac{1}{2} s(e)\epsilon^2) \leq 0. \tag{1.5}$$

— *The optimal recognition effort e increases with the reward δ. Alternatively, there exists a threshold level of permissiveness $\bar{\epsilon}(\delta)$ that decreases with δ such that when $\epsilon < \bar{\epsilon}(\delta)$, more permissiveness increases the child's recognition effort (i.e. e increases with ϵ). Alternatively, when $\epsilon > \bar{\epsilon}(\delta)$, permissiveness trumps out the child's recognition effort (i.e. e decreases with ϵ).*

By investing in improving the quality of his signal, the child knows that he increases his probability of choosing an action that lies in the interval $[a - \epsilon, a + \epsilon]$. The marginal benefit of increasing e is the second term in (3.5) (this result obtains after few computations given in the Appendix). The second-order condition is necessarily verified so there is a unique optimal strategy for the child.

It is straightforward from (3.5) that the marginal benefit of increasing the quality of the signal is increasing in the magnitude of the reward δ. Alternatively, a higher degree of permissiveness ϵ has an ambiguous effect. The intuition is represented in figure 3.1. The curve

above area C represents the density $g(.,e+de)$ for $de>0$ while the other one represents $g(.,e)$. On the one hand, a higher quality of the signal pushes upward the probability that the actions in the neighborhood of the mean are close to the parent's optimum a. But on the other hand, the actions that are in the neighborhood of the thresholds $\overline{a}\pm\epsilon$ become less likely to be close to the action expected by the parent. Thus, if ϵ is sufficiently high, then when ϵ decreases, the distribution becomes more concentrated around the mean and the loss from a higher precision (area $A+B$) may be higher than the benefit of a higher precision (area C). In other words, if permissiveness is high, the child prefers to over-estimate his probability of receiving the reward and thus chooses to decrease his recognition effort. An increase in the precision of the signal decreases the welfare of the child because it shows that the probability of being wrong is higher than previously anticipated. Alternatively, a small ϵ means that the child gets the reward in the close neighborhood of the mean, so he always benefits from a higher precision.

The parent's decision problem can be written as

$$\max_{\delta,\epsilon,c_p} \mathbb{E}_{a^{obs}} W = \alpha v_c(e(\delta,\epsilon)) - \lambda_p \mathbb{E}_{a^{obs}}[\mathbb{E}_{a^*}(a^*-\overline{a})^2 \mid a^{obs},a] + u(c_p), \tag{1.6}$$

given the constraint $c_p + P\delta \leq y_p$. The parameter P is either equal to 1 when the parent expresses love and 0 otherwise. We will assume in the sequel that the parent's degree of altruism toward her child α is sufficient large, so that she always rewards him in equilibrium and $P=1$. The parameter c_p relates to the parent's private consumption during the rearing

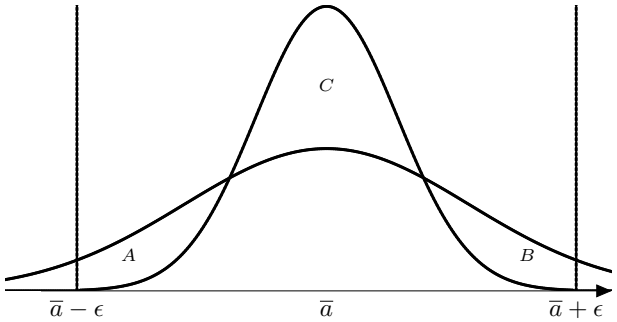

Figure 1.1 – Recognition effort and rearing strategy

period and y_p to his income. We label the parameter $\lambda_p \geq 0$ the degree of paternalism in child rearing while $\alpha > 0$ is the degree of altruism. The first term in (3.6) relates to the pure altruistic motive in inter-generational transmission, as $v_c(e(\delta, \epsilon))$ is the present utility of the child given that he invests an effort $e(\delta, \epsilon)$ in recognition that solves (3.5). The second term in (3.6) gives a quadratic loss - felt by the parent - from her offspring choosing action $\overline{a}$ instead of a^*. The expectation factors correspond (i) to the uncertainty over the signal a^{obs} and (ii) to the uncertainty over a^* evaluated by the parent, i.e. given her prior on the distribution of a^* (recall that a^* is the correct action to perform for the child, as evaluated by her parent). Since the parent believes that the best action a^* is drawn from a normal distribution with mean a and precision s_p, we deduce that

$$\mathbb{E}_{a^{obs}}[\mathbb{E}_{a^*}(a^* - \overline{a})^2 \mid a^{obs}, a] = z(e) = \frac{1}{s_p} + \frac{1}{(s_0 + f(e))^2}\{f(e) + s_0^2(a - a_0)^2\}. \tag{1.7}$$

If the parent was acting instead of the child, she will choose action a and thus her quadratic loss will reduce to $1/s_p$. Here, since the child has a posterior belief that is biased relative to the prior distribution of the parent, there is a bias on top of the parent's variance $1/s_p$ in (3.7). Since $s(e) = s_0 + f(e)$ increases with e, we can show that the quadratic loss in (3.7) is lower the higher is the precision e over the segment $[0, \infty)$ whenever the condition $s_0(a - a_0)^2 > 1/2$ is fulfilled. [8] Intuitively, $a - a_0$ must be large enough so that the signal a^{obs} makes the posterior of the child converge toward a. If the two parameters are too close, then the signal may induce a divergence of the child's posterior and the parent's optimum.

Assumption 1. *The child and the parent's prior a_0 and a respectively are sufficiently apart from each other, $s_0(a - a_0)^2 > 1/2$.*

A strictly positive value of λ_p may relate to the transmission of a behavior which consequences are not directly foreseen by the child in his infancy but that may affect his behavior later on. Indeed, a parent may transmit to delay consumption in some specific contexts

8. By differentiating the second term in the RHS of (3.7), we find that its derivative with respect to e is negative whenever $s_0 - f(e) - 2s_0^2(a - a_0)^2 < 0$. Consequently, if the preceding inequality holds for $e = 0$, then it is valid for any positive value of e. We deduce that the RHS of (3.7) decreases with e when $1/2 < s_0(a - a_0)^2$, as stated in Assumption 1.

and such behaviors may then allow for the development of behavioral norms in the child's knowledge structure that are congruent with the norm patience in his adult life.

The parameter P is the probability that the child receives the reward δ from his parent. Indeed, if the parent does not care about the child's cognitive development (i.e. $\alpha = 0$ and $\lambda_p = 0$), then there is no point in investing in transmitting anything so the parent will simply use his income for his private consumption, $c_p = y_p$. Similarly, from the moment that $\overline{a}$ does not reach the interval $[a \pm \epsilon]$, then the reward is not given and the budget constraint simplifies to $c_p \leq y_p$. We will abstract from this issue in the rest of this section by assuming that the parent cares sufficiently for her offspring so that the latter is necessarily rewarded in equilibrium.

We obtain the first-order conditions associated with the optimization program of the parent with respect to δ and ϵ by substituting $\overline{a}$ with (3.1) and $v^c(e(\delta, \epsilon))$ with (3.5). The first-order conditions associated with the optimization of the parent writes

$$\frac{\partial \, \mathbb{E}_{a^{obs}} W}{\partial \delta} = \alpha \int_{\overline{a}-\epsilon}^{\overline{a}+\epsilon} g(u, e) du + \frac{\partial e}{\partial \delta} z'(e) - u'(y - \delta) = 0, \tag{1.8}$$

and

$$\frac{\partial \, \mathbb{E}_{a^{obs}} W}{\partial \epsilon} = \alpha \delta g(\overline{a} + \epsilon) + \frac{\partial e}{\partial \epsilon} z'(e) = 0 \tag{1.9}$$

when the solution is interior.

First, increasing the reward δ has a direct effect on the utility of the child, so it feeds back into the parent's utility when $\alpha > 0$. This is the first term in the LHS of (3.8). Second, an increase in δ increases the effort invested by the child in improving the quality of the signal he receives during the interaction with his caregiver. Thus, the child adopts an action that is closer from the optimum of the parent, which reduces the quadratic loss associated with the paternalist motive of the latter. Finally, the marginal cost of increasing the reward is equal to the marginal benefit of private consumption for the parent.

Relative to the choice of ϵ, the parent has a trade-off between the direct benefit of a higher permissiveness on the utility of the child (first term) and the effect of a higher permissiveness

on the child's recognition effort (second term). Indeed, too much permissiveness decreases the incentive of the child to invest in recognition along the lines of Proposition 1. In turn, this will increase the quadratic loss of the parent, because the action of the child $\overline{a}$ will diverge from a.

Proposition 2.

- *If the parent is altruistic (i.e. $\alpha > 0$), then she is necessarily too much permissive in equilibrium as the optimal level of permissiveness trumps out the child's recognition effort, $\partial e/\partial \epsilon < 0$.*
- *There may be several equilibria. If α/λ_p is sufficiently high, there exists a complementarity between a higher permissiveness ϵ and the magnitude of the reward δ along the equilibrium path. Thus, given that α/λ_p is sufficiently high, a poorer parent is predicted to be more authoritarian while her child is predicted to invest less in the rearing interaction, ceteris paribus.*
- *A higher degree of paternalism λ_p makes the parental rearing strategy more rewarding for the child (i.e. δ increases), more authoritarian (i.e. ϵ decreases) and increases the child's investment in recognition. A higher degree of altruism α has an ambiguous effect on the rearing strategy and on the child's recognition effort.*

Observe first that when $\lambda_p = 0$, there is no cost associated with being more permissive, because the parent does not care about the congruence between his optimum and the action chosen by her child. Consequently, purely altruistic parents are excessively permissive and $\epsilon \to \infty$. Furthermore, high levels of permissiveness have dramatic consequences on the child's investment in state recognition, since $\partial e/\partial \epsilon < 0$ in equilibrium. Thus, we should expect excessively permissive parenting styles to correlate with lower cognitive development outcomes. Existing evidence are consistent with this prediction. For instance, Piotrowski, Lapierre, and Linebarger (2013) find that parents who have notable absence of control are more likely to have children with considerable regulatory deficits (the children age between 2 and 8 years old in the data of the previous authors). Interestingly, Piotrowski, Lapierre, and Linebarger (2013) also find that parents that are children of excessively authoritarian parents will develop significant regulatory deficits. This accords as well with our prediction that when ϵ is excessively low,

then so should be the child's investments in state recognition. Similarly, using data from the United Kingdom, Chan and Koo (2011) find that compared to adolescent with authoritative parents, education achievements are significantly lower for those with either permissive or authoritarian parents.

Second, from (2.26), it must be that $\partial e/\partial \epsilon < 0$ in equilibrium, meaning that the parent is too much permissive. Take a low initial value of ϵ. The parent has an incentive to increase his degree of permissiveness for two reasons. First, if she is altruistic (i.e. $\alpha > 0$), then increasing ϵ implies rewarding the child more often so it increases her utility. Second, the parent knows that increasing ϵ will affect positively the recognition effort invested by her child for low values of ϵ, i.e. for $\epsilon < \bar{\epsilon}(\delta)$, see Proposition 2. Thus, the degree of permissiveness is necessarily above $\bar{\epsilon}(\delta)$ and the parent equalizes the marginal benefit of rewarding the child more often (first term in (2.26)) with the marginal loss implied by a decreasing precision of the signal received by the child during the rearing interaction (second term in (2.26)).

We establish in the Appendix by differentiating (3.8) with respect to ϵ that there exists a complementarity between permissiveness and reward along the equilibrium path when α/λ_p is sufficiently high. Consider (3.8). Increasing the level of permissiveness ϵ will increase the likelihood that the child is rewarded, so it increases his utility and thus the marginal benefit at giving pecuniary benefits to the child when the parent is altruistic. Altruism creates a first motive for the complementarity between permissiveness and reward. Additionally, observe that a higher value of ϵ will decrease the precision of the signal in equilibrium, so this will tend as well to increase the marginal benefit at investing resources in child rearing, given that the parent is paternalist (second term in (3.8)). Those two effects explains the existence of the complementarity between pecuniary investments and permissiveness in child rearing. There exists nevertheless a third effect that creates an ambiguity in the general case since a higher ϵ makes it harder to increase the precision of the signal by investing in the reward δ.

Whenever α/λ_p is sufficiently high, a higher income increases the magnitude of the reward δ while - by complementarity - it raises the level of permissiveness. Thus, on the one hand, a poorer parent will invest less in child rearing, i.e. δ will decrease, and this will decrease the

incentive of the child to invest in recognition. On the other hand however, a poorer parent will be more authoritarian, i.e. ϵ will decrease, and this will tend to increase the child's propensity to invest in recognition. In any case, the higher degree of authoritarianism will not compensate entirely for the lower pecuniary investment in equilibrium by a revealed preference argument. Indeed, the utility of the parent is unambiguously increasing in the effort $e(\delta, \epsilon)$ invested by the child during the rearing interaction. Thus, if a rich parent - by adopting the strategy of a poorer parent, which is feasible for him given his higher income - could incentivize the child to make a higher recognition effort, she will do so. Thus, a poorer parent that invests less in the rearing interaction necessarily incentivizes less her child to invest in recognition.

According to the theory, we should expect poorer parents to adopt relatively more authoritarian parenting styles. Although evidence on the subject are scarce, the existing studies find that poorer parents tend to rely more often on excessively authoritarian rearing methods such as corporal punishments, see for example Straus and Stewart (1999) and Weinberg (2001).

Furthermore, our results are also consistent with series of evidence showing that there exist positive linkages between parental income and various measures of children cognitive development. Using the Baylay Scales of Infant and Toddler Development on a sample of children aged 6-42 months in Bogota, Rubio-Codina, Attanasio, Meghir, Varela, and Grantham-McGregor (2015) find an average difference of 0.53, 0.42 and 0.49 standard deviations in cognition, receptive and expressive language respectively between children in the top and bottom quartile of the wealth distribution. Similar results are obtained by Schady, Behrman, Araujo, Azuero, Bernal, Bravo, Lopez-Boo, Macours, Marshall, Paxson, and Vakis (2015) and Fernald, Kariger, Hidrobo, and Gertler (2012) for example, among others. Additionally, Hair, Hanson, Wolfe, and Pollak (2015) demonstrated with data of imaging scans of 389 children that those living 1.5 times below the federal poverty level had smaller volumes of several brain regions critical for cognitive and academic performance (see as well Luby (2015)). They show that children from low-income households scored 4 to 7 points lower on standardized tests of academic achievements, and that 20% on the gap in test scores could be explained by

maturational lags in the frontal and temporal lobes.

The comparative statics given in the third point of the Proposition are intuitive. Whenever the degree of paternalism λ increases, the parent has a higher marginal benefit at increasing the precision of the recognition of his offspring, so she increases the reward δ and reduces the degree of permissiveness ϵ. Paternalism increases the propensity to rely on authoritarian rearing strategies. This accords with the analysis of Doepke and Zilibotti (2014), who show that in traditional societies with a strong incumbency advantage and low social and occupational mobility - i.e. societies where paternalistic motives in transmission are high - we should expect authoritarian parenting styles to dominate. Alternatively, when the parent is more altruistic (i.e. α increases), then the parent rewards more (i.e. δ increases) and more often (i.e. ϵ increases). Thus, "too much love" in the sense of a high valuation of the present utility of the child is not associated with the best cognitive development outcomes for the child, since higher levels of permissiveness trump out his investment in recognition.

1.2.2 Television viewing for rearing purposes

There has been an explosion of videos and television viewing in the lives of infants and toddlers in recent years. The study of the effects of those media technologies on children development and parents' rearing practices is nevertheless still in its infancy, although studies on the first subject have been flourishing in recent years in the fields of pediatric and developmental psychology.

The American Academy of Pediatrics (AAP) recommends no screen time for children younger than 2 years and only high-quality, age appropriate viewing thereafter (The American Academy of Pediatrics (2001)). Although no scientific evidence was provided by the committee, as argued by Courage and Howe (2010), the implication was that the time spent viewing video was not spent playing, interacting socially or acquiring language skills, which are essential for early brain growth and cognitive development. The recommendation of the AAP is largely ignored in the US, since 32% of parents with children younger than 2 years of age are complying according to Certain and Kahn (2002). More recently, Zimmerman, Christakis,

and Meltzoff (2007a) estimated that by 3 months of age, about 40% of children regularly watched television, DVDs, or videos. By 24 months, this proportion rose to 90%.

Although the literature has not reached a consensus on the impact of video materials on infants and toddlers, some studies find significant negative effects on cognitive outcomes. For instance, Zimmerman and Christakis (2005); Zimmerman, Christakis, and Meltzoff (2007b); Christakis, Zimmerman, DiGiuseppe, and McCarty (2004) produced evidence from large-scale surveys (e.g. National Longitudinal Survey of Children and Youth), that infants and toddlers who are heavy television viewers are at risk for deficits in attention and perform more poorly on the MacArthur-Bates Communicative Development Inventory. [9]

Similarly, studies of the co-determinants of infants and toddlers television viewing are still largely missing, although existing evidence points toward negative effects of parents' education and household income on infants and toddlers television viewing. Using data from telephone surveys conducted in 2005 in the US of about 1,000 parents of children ages 6 months to 6 years old, Rideout and Hamel (2006) find that children from families with incomes under $20,000 a year spend on average of almost half an hour more watching television each day than do children from families earning $75,000 a year and up. Similar differential patterns occur between children whose parents have a high school education or less, as compared to children with a parent who is a college graduate. Furthermore, numerous surveys show that parents hold the pervasive beliefs that age-appropriate videos can have important positive educational effects and provide good entertainment for the children (Zimmerman, Christakis, and Meltzoff (2007b), Rideout and Hamel (2006)). Finally, parents usually report as well using the television to substitute for their own involvement in child rearing (Zimmerman, Christakis, and Meltzoff (2007b), Rideout and Hamel (2006)).

In order to study the determinants of infants and toddlers television viewing as well as its effects on child rearing and cognitive outcomes, we extend the model of the previous section. As before, we focus on the transmission of a single state $w \in \Omega$ that belongs to the knowledge

9. Other analyses have failed to replicate the previous findings and find weak or non-existing effects of television on cognitive outcomes, see for instance Foster and Watkins (2010) and Obel, Henriksen, Dalsgaard, Linnet, Skajaa, Thomsen, and Olsen (2004).

structure of the parent and that is associated to an action a^* that is - from the parent's perspective - drawn from a normal distribution with mean a and precision s_p. During the child-caregiver interaction, the parent still transmits a signal to the child on the best action to perform a^{obs} that is drawn from a normal distribution centered on a, with a precision $f(e)$ endogenous to the child's effort in the interaction. We assume that during the child-rearing interaction, the parent decides the duration of television viewing t. Watching television is costless for the child and transmits him a signal b^{obs} on the optimum action that is drawn from a normal distribution with mean b and precision $g(t)$. We assume $g(.)$ increasing and concave with $g(0) = 0$. The parent knows $g(t)$ and b.

If $b \equiv a$ for example, then watching the television allows to the child to have a more precise understanding of the action that is preferred by his parent. We could imagine that by wisely choosing the program watched by her child, the parent could make b as close to a as possible. However, several characteristics inherent to the "child-television" interaction make it clearly different from "real life" interactions with a caregiver. Such characteristics include for instance the mere format of television, i.e. a two dimensional space and the disrespect of the laws of physics in several television programs for infants and toddlers. Furthermore, television presents the viewer with a more impoverished visual and auditory array than the real world (Anderson and Hanson (2010)). The image occupies much less of the viewer visual field and provides a more limited selection of cues for the perception of depth. As argued by Anderson and Hanson (2010), television viewing also requires the individual to understand the codes and conventions that characterize television as a medium of communication. However, infants and toddlers may not have the cognitive ability to understand those codes and conventions before 18 months of age at least. We therefore assume that there exists a discrepancy between a and b with $b = a - \eta$ with $\eta > 0$ so $a_0 < b < a$.

Although it provides slightly biased signals, the television increases the child's understanding by granting - ex post - a higher precision to his beliefs on the best possible action. Indeed, the posterior distribution of the child on the correct action to perform is normally

distributed with a precision $s = s_0 + f(e) + g(t)$ and a mean

$$\overline{a} = \frac{a_0 s_0 + a^{obs} f(e) + b^{obs} g(t)}{s_0 + f(e) + g(t)}. \tag{1.10}$$

As in the previous section, the child chooses his effort e so as to maximize

$$v^c(e) = \delta \int_{-\epsilon}^{\epsilon} \sqrt{\frac{s}{2\pi}} \exp(-\frac{1}{2} s u^2) du - \frac{e^2}{2}, \tag{1.11}$$

with $s = s_0 + f(e) + g(t)$. Differentiating the first-order condition with respect to t gives

$$\frac{\partial e}{\partial t} = -\frac{g'(t)}{f'(e)}, \tag{1.12}$$

meaning that the child substitutes his effort in the rearing interaction with television viewing. The preceding equation says that if the child watches the television one more unit of time, then the precision of his information increases by $g'(t)$ and he relaxes his effort in the child-parent interaction so as to keep the precision of his signal constant. A positive time of television watching decreases the need of being involved with the parent since it is a free source of knowledge acquisition that competes with it. This result will be relatively similar if the child had a small cost from watching the television, say to stay focused for example. Our main point is that if we think of the acquisition of knowledge - and of brain development for that matter - as affected by several channels of interaction (e.g. a television and a caregiver), then we should expect the child to optimize his recognition effort across those different channels. Such an optimization implies an equalization of the marginal returns of the available sources of knowledge.

Furthermore, observe that if the time of television viewing increases, then $\overline{a}$ necessarily converges toward b, not toward a the optimum of the parent. This is because the child substitutes entirely his effort in the rearing interaction with television viewing, yet he does not foresee that television gives him a biased signal on the action that is expected by the parent. Hence the need for the latter to control the time of television viewing. The parent's

decision problem can be written as

$$\max_{t,\delta,\epsilon,c_p} \mathbb{E}_{b^{obs}} \mathbb{E}_{a^{obs}} W = \alpha v_c(e(t,\delta,\epsilon)) - \lambda_p \mathbb{E}_{b^{obs}} \mathbb{E}_{a^{obs}} [\mathbb{E}_{a^*}(a^* - \overline{a})^2 \mid a^{obs}, b^{obs}, a] + u(c_p), \quad (1.13)$$

given the constraints $c_p + \delta \leq y_p$ (we assume, as before, that the child is always rewarded in equilibrium) and $0 \leq t$. Since the parent believes that the best action a^* is drawn from a normal distribution with mean a and precision s_p, we find that

$$\mathbb{E}_{b^{obs}} \mathbb{E}_{a^{obs}} [\mathbb{E}_{a^*}(a^* - \overline{a})^2 \mid a^{obs}, a] = w(e,t) = \frac{1}{s_P} + \frac{1}{(s_0 + f(e) + g(t))^2} \{f(e) + g(t) + s_0^2 (a - a_0)^2 + 2\eta g(t) s_0 (a - a_0) + \eta^2 g(t)^2\}. \quad (1.14)$$

Observe then that when $\eta = 0$, we find (3.7) with a precision $f(e)+g(t)$ instead of $f(e)$. Indeed when $\eta = 0$, it is as if the child had access to two signals centered on a sent simultaneously by the parent with precisions $f(e)$ and $g(t)$. The existence of a discrepancy between the signal sent by the television and that sent by the parent creates a positive bias in the quadratic loss that is given in the second line of (1.14). Furthermore, since e depends on t in equilibrium and

$$f'(e) \frac{\partial e}{\partial t} + g'(t) = 0, \quad (1.15)$$

it is direct that the LHS of (1.14) is increasing with t for $\eta > 0$, meaning that letting the child watch the television necessarily increases the quadratic loss of the parent. Thus, when the parent chooses the time of television viewing t, she trades-off the increase in the welfare of the child induced by the television with her own quadratic loss. The first-order condition writes

$$\alpha \frac{f'(e) e}{g'(t)} - \lambda_p \frac{dw(e(t,\delta,\epsilon),t)}{dt} \leq 0, \quad (1.16)$$

and holds with equality if $t > 0$.

Thus, when the parent is sufficiently altruistic (i.e. α/λ_p is sufficiently high), she lets her offspring watch the television for a positive period of time t in equilibrium. If the parent does

not care about the welfare of the child when young (i.e. $\alpha = 0$ and $\lambda_p > 0$), then she does not allow for a positive time of television watching because she knows that the television will make her child's beliefs diverge from a.

We can now study how the television affects the rearing strategy (δ, ϵ). First, the effect of television on the magnitude of the pecuniary reward δ is positive *ceteris paribus*. This is because a higher time of television watching makes the child deviate from a. In order to counter this effect, the parent needs to invest more resources in child rearing. A higher time of television watching should make the parent less permissive for the same reason, *ceteris paribus*. Thus, when ϵ and δ are complement along the equilibrium path, then the television has an ambiguous effect on the rearing strategy. Indeed, the direct effect of t should be to increase δ and to decrease ϵ, i.e. it should make the parent more involved in child rearing. But the indirect effect here matters as well, since if δ and ϵ are complement, the potential increase in δ that follows a higher time of television watching mitigates the negative effect of t on ϵ. If the indirect effect dominates and the effect of television on the parent's investment is strong, then it could be that a higher time of television watching makes the parent more permissive while it increases her investments. Similarly, if the effect of television watching on the degree of authoritarianism is strong, then we should expect that an increase in t reduces so much ϵ that it reduces as well the parent's investment δ in child rearing.

The model is too general to make predictive statements on the way television viewing affects the rearing strategy, although it makes it clear that infants' television watching leads to lower investments in state recognition in "real-life" rearing interactions. Furthermore, when the indirect effects dominate, television viewing may lead to inefficient rearing strategies. When the income of the parent is low for example, then an indirect effect is likely to be dominant because the parent can not adjust her investment in child rearing δ to the duration of television viewing. Thus, a poorer parent will take advantage of the television in order to improve costlessly the welfare of her child when she is altruistic. She will yet not be able to compensate the child's television viewing by higher investments in child rearing given her income constraint. This result accords with the study of Rideout and Hamel (2006) for

example that points toward positive correlations between lower income at the household level and higher time of TV watching from young children.

1.3 The dynamics of child rearing and child development

We have been concerned so far with the transmission of a single state of the world w from a parent to her child. Doing so allowed us to introduce a simple mechanism of inter-generational transmission and to study how the parent's characteristics (e.g. income and various preference parameters) affect her degree of authoritarianism as well as her pecuniary investments in child rearing. However, such a static theory can not answer for the emergence of behavioral norms through early rearing interactions. In this section, we intend to fill this gap by studying the formation of a cognitive capital through repeated child rearing interactions aiming at transmitting norms.

Definition 1. *Any agent that possesses a behavioral norm A recognizes a series of states of the world $\Omega^A = \{w_1^A, \ldots, w_p^A\}$ where norm A prescribes a specific collection of behaviors $a(\Omega^A) = \{a(w_1^A), \ldots, a(w_p^A)\}$. Any norm has an internal consistency, i.e. there exists a common perception of the world associated with the states in Ω^A.*

A behavioral norm is associated with both a perception of different states of the world and contingent action rules (i.e. $a(\Omega^A)$). Take the example of patience. This norm - broadly defined - relates to the delaying a present activities, say consumption, given that one foresees the long-term benefits of the delay. Foreseeing the future stream of benefits is the perception of the world associated with the norm patience. This perception applies for a plethora of situations (i.e. states of the world). There exists empirical evidence that shows that few norms guide human behaviors. Almlund, Duckworth, Heckman, and Kautz (2011) for instance show that stable character skills exist and are predictive of many behaviors. Similarly, analyses of the World Value Surveys, e.g. Inglehart and Welzel (2010), reveal two major dimensions

of cross-cultural variations in beliefs and values, (i) a traditional versus secular-rational dimension and (ii) a survival versus self-expression dimension.[10] Besides patience, important behavioral norms that are taught early in life include self-control, trust, attentiveness or self-esteem (Heckman and Kautz (2013)).

In order to explain how child rearing creates endogenous preferences, we rely on a powerful idea in the literature that can be traced back at least to Gary S. Becker and co-authors. It says that preferences can be understood as a capital in the standard sense used in the economic literature and that the agents build progressively a taste for this capital. In their study of addictive behaviors for instance, Becker and Murphy (1988) writes "[...] smoking of cigarettes, drinking of alcohol, injection of heroin, or close contact with some persons over an appreciable period of time, often increases the desire (creates a craving) for these goods or persons, and thereby cause their consumption to grow over time". Similarly, one tends to appreciate being patient as one acts in a way that is consistent with this norm.

Behaving consistently with any behavioral norm requires understanding that a given situation corresponds to a state of the world where the norm dictates a specific behavior. This observation is key because it means that any behavioral norm has to be supported by a perception of the world, e.g. foreseeing long-term consequences of delaying consumption for patience. This, in turn, reveals the importance of early rearing in the acquisition of behavioral norms since parents' early investments in child rearing mold their children perception of many states and thus support the emergence of an appreciation capital that conditions the perception of the child's future interactions.

We model the formation of a cognitive capital in a dynamic principal-agent framework where the principal - the parent - can reward her offspring for behaving a certain way in different situations, knowing that "good" behaviors allow the child to create an appreciation for the norm. We therefore extend the static model of the previous section and assume that at time t, the state of the world w_t^A is taught by the parent with A the behavioral norm that dictates a behavior $a(w_t^A)$. We focus on the behavioral norm A in the sequel, so we omit the

10. Less recently, Epstein (1980) presented evidence that people act in a predictable fashion with a high level of reliability of average behaviors.

superscript A (i.e. we assume that $\Omega \equiv \Omega^A$).

Adapting the model of Becker and Stigler (1977), we assume that the utility of the offspring in a given period t depends on M_t, his "appreciation" of the behavioral norm A at time t. Furthermore, M_t depends on both a cultural capital accumulated until period t that we denote S_{t-1} and the effort e_t invested by the child in period t:

$$M_t = u(e_t, S_{t-1}), \tag{1.17}$$

with $u(.,.)$ a concave function such that

$$\frac{\partial u}{\partial e_t} > 0, \frac{\partial u}{\partial S_{t-1}} > 0 \text{ and } \frac{\partial^2 u}{\partial e_t \partial S_{t-1}} > 0. \tag{1.18}$$

A marginal increase in the cognitive capital tends to increase the child's marginal utility at increasing his recognition effort. We posit that S_{t-1} is a weighted function of the recognition efforts invested by the child in learning norm A until period t,

$$S_{t-1} = \sum_{i=0}^{t-1} z_{i,t} e_i, \tag{1.19}$$

where $z_{i,t} \geq 0$ gives the weight of the state of the world taught at time i w_i in the building of a capital of appreciation in norm A at time t. It can be for instance that $z_{i,t} = d^{t-i} h_i$ with d a rate of depreciation per period that is common to all the traits of the knowledge structure and h_i a state-specific contribution. We posit that $d \equiv 1$ and $h_i \equiv 1$ for simplicity of exposition in the sequel, meaning that there is no time-depreciation and all the states of the world in Ω^A contribute equally to the building of the norm.

We assume that in each period, the timing of the game is similar to that given in the static model. During the rearing interaction, the child still receives a signal, which precision is endogenous to his investment in state recognition. The signal is not necessarily sent by the caregiver. It can relate to an interaction with another human being, say a teacher or a family member, or with an object like a television. The parent decides at the begining of each

period if she wants to supervise her child's interaction. By supervising the interaction, the parent can help the child through a pecuniary and conditional reward scheme to interpret the situation he is in. For simplicity then, we assume that when the parent supervises the interaction, it is as if she was sending herself the signal received by the child in his interaction. Thus, we are back to the precise settings of the static model of the previous section when the parent mediates the interaction. If she does not, then the child behaves independently given the cognitive capital he has accumulated so far.

In each period, the offspring maximizes his present utility, which now depends on both the strategy of his parent - when she supervises the interaction - and his own appreciation of behavioral norm A at time t,

$$v^c(e_t) = \delta_t \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t u^2) du - c(e_t) + u(e_t, S_{t-1}), \tag{1.20}$$

with $s_t = s_0 + f(e_t)$. As before - and given the concavity assumption on $u(.,.)$, we deduce that there is a unique solution $e(\delta_t, \epsilon_t, S_{t-1})$ to the child's maximization problem in period t. The comparative statics of the previous section are robust. The only novelty is that the recognition effort in period t depends positively on the capital accumulated until period t, meaning that the higher the cognitive capital, the higher the effort invested by the child in behaving according to the norm. Thus, given that the cognitive capital at time t S_{t-1} is increasing with any e_τ for $\tau \leq t-1$, there is a dynamic complementarity in the offspring's recognition effort. As there is a common perception associated with the states belonging to norm A, when facing a situation that potentially corresponds to the norm ex ante, the child is incentivized by his cognitive capital in the norm to invest in recognizing more precisely the situation he is involved in. Cunha and Heckman (2007) and Cunha, Heckman, and Lochner (2006) for example account as well for complementarities in cognitive development, which they assume are characteristics of the child's production function of skills. Our point here is that complementarities arise because skills participate to the accumulation of an appreciation capital.

We now turn to the optimization program of the parent. We assume that the parent can be active in inter-generational rearing for at most T periods. Her utility writes

$$W(0) = \sum_{t=0}^{T} \beta^t \{\alpha v^c(e(\delta_t, \epsilon_t, S_{t-1})) - \lambda z(e(\delta_t, \epsilon_t, S_{t-1})) + u(c_t)\}, \tag{1.21}$$

with

$$\sum_{t=0}^{T} \frac{1}{(1+r)^t}\{c_t + P_t\delta_t\} \leq \sum_{t=0}^{T} \frac{y_t}{(1+r)^t} \equiv Y \tag{1.22}$$

given an initial capital $S_{-1} \equiv 0$, a revenue Y on the rearing period, an interest rate $0 < r < 1$ and a discount factor $0 < \beta < 1$. $P_t = 1$ if the reward is given to the child in period t and zero otherwise. $z(.)$ is given in (3.7). We will denote $e(\delta_t, \epsilon_t, S_{t-1}) \equiv e_t$ and $s_t \equiv s_0 + f(e_t)$ hereafter. The optimal paths of δ_t and ϵ_t are determined by the first-order conditions

$$\alpha \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2}s_t u^2)du + \frac{\partial e_t}{\partial \delta_t}\{-\lambda_p z'(e_t) + A_t\} = \frac{\mu}{(\beta(1+r))^t} \tag{1.23}$$

and

$$2\alpha\epsilon_t\sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2}s_t\epsilon_t^2) + \frac{\partial e_t}{\partial \epsilon_t}\{-\lambda_p z'(e_t) + A_t\} = 0, \tag{1.24}$$

with

$$A_t = \sum_{\tau=t+1}^{T} \beta^{\tau-t}\{\alpha\frac{\partial u(e_\tau, S_{\tau-1})}{\partial S_{\tau-1}} - \lambda_p z'(e_\tau)\}\frac{\partial S_{\tau-1}}{\partial e_t} \tag{1.25}$$

and $\mu \geq 0$ the Lagrangian coefficient associated with the constraint (3.17). The expression A_t represents the effect of an increase in the recognition effort in period t on the future stock of cognitive capital. There are two effects to distinguish in A_t. First, an increase in e_t will affect the cognitive capital in subsequent periods *ceteris paribus*. This, in turn, will affect positively the utility of the child and thus will feedback into the utility of the parent when she is altruistic (i.e. when $\alpha > 0$). Second, a higher value of e_t - by increasing the cognitive capital - will affect the propensity of the child to invest in recognition in future periods. This effect stems from the dynamic complementarity between e_t and e_τ for $\tau > t$. Thus, if the parent is paternalist (i.e. $\lambda_p > 0$), then she will anticipate that early investments increase the

future recognition efforts.

We can reasonably expect A_t to be higher in early periods and fades as t approaches T. The parent's incentive to invest early in child rearing may then be particularly high because early investments initiate the emergence of a cognitive capital that begets' the child's future investments in recognition. Furthermore, from (24), we should expect the parent to be more authoritarian in early periods for similar reasons.

Observe that pure altruism (i.e. $\alpha > 0$ and $\lambda_p = 0$) appears as less problematic relative to the static model since it does not necessarily promote a high permissiveness level and a limited investment of the child in state recognition. Recall that whenever $\lambda_p = 0$ in the static model, then the parent is completely permissive, i.e. $\epsilon \to \infty$ and the child's effort in recognition is low since he always gets rewarded for what he does. This result is not robust to the dynamic extension presented above because the altruistic parent now anticipates that being too much permissive deteriorates the formation of a stock of capital, which he cares about as well. Going a bit further, this result means that pure altruism is not antithetical with the persistence of cultural heterogeneity, since it does not preclude the efficient transmission of an appreciation capital.

As the child builds his cognitive skills - through the supervision of his parent - he becomes more independent in recognizing the states of the world where norm A applies. At some point, the parent may stop mediating her child's interactions since the latter have accumulated enough cognitive capital to act consistently with the norm on his own. This outcome will obtain when the incentive compatibility constraint on positive investments in child rearing binds and the parent then simply trades-off her own present and future consumptions (given her discount factor β and the prevailing interest rate r). Observe that this outcome obtains when the dynamic complementarity is strong enough. When $f(.)$ is not bounded above for instance, T is sufficiently high and Ω^A contains an arbitrarily high number of states for example, then the parent almost surely stops supervising her child's interaction after some period $t < T$ and the child then chooses behaviors that exactly correspond to the behavioral

norm of his parent after some period t_1, with $t_1 \geq t$.[11]

It is interesting to note that a parent that transmits during the whole rearing period T does not necessarily produce more cognitively competent children, despite her investments in affection being possibly significantly high over the rearing period. Take the example of a parent with a low discount factor β for example. This parent fails to recognize the effect of her present level of affeection on the formation of a stock of cognitive capital. Thus, she may be steadily active in transmitting a behavioral norm although with weaker early investments in affection relative to a parent with a higher β. This will make the offspring invest low levels in recognition throughout the rearing period and thus he may never develop enough cognitive capital to be able to interpret the signals he receives on his own.

The discount factor of the parent plays a central role in explaining her ability to efficiently transmit a cognitive capital. The key mechanism here is that a lower discount rate favors a steeper investment profile that takes advantage of the dynamic complementarity in the child's investments in recognition. As the parent foresees the importance of a cognitive capital in motivating the future investments of her child, she invests more in early rearing periods. Series of evidence shows that parents with a higher human capital produce more cognitively competent children and have a higher ability to transmit their behavioral norms. For instance, Schady (2011) shows that in a longitudinal study of relatively poor children in Ecuador, the schooling and vocabulary levels of mothers were strong predictors of the cognitive development of the young children. Similarly, Rubio-Codina, Attanasio, and Grantham-McGregor (2016) show that more educated mothers provided better home stimulation than less educated ones, with positive effects on children cognitive development. With the preceding theoretical foundations, what appears as critical is the higher time preferences that correlate with higher levels of human capital. Indeed, if human capital is reduced to its positive effect on economic status for example, or is accounted for through a more efficient technology of state recognition for the child, then the effects on the child's investments in recognition may be ambiguous at best since the level of permissiveness will increase along the equilibrium path (see Proposition

11. $t_1 \geq t$ because the parent may stop being involved whenever the actions of the child are close to their optimum, but not "arbitrarily" close.

3).

An interesting research has recently tried to identify the effects of parental beliefs on the nature of the production function of children human capital. For instance, Rowe (2008) argues that gaps in the early home language environment exist because poor, uneducated mothers do not know about the role they play in determining the language and cognitive development of their children. This, in turn, may be explained by the lower time horizon of uneducated women, since it is precisely what makes them underestimate their influence in their children accumulation of skills. Similarly, Aizer and Stroud (2010) have tracked the smoking habits of educated and uneducated pregnant women before and after the release of the 1964 Surgeon General's Report on Smoking and Health. Before the release of the report, educated and uneducated pregnant women smoked at roughly the same rates. After the report however, the smoking habits of educated women decreased and a ten percentage point gap in pregnancy smoking rates between educated and uneducated women was observed. Again - and consistently with both the seminal theory of Becker and Murphy (1988) on addictive behaviors and the closely related model of this paper - the difference in smoking behaviors of educated and uneducated pregnant women may be explained by differences in time horizons. Finally, Cunha, Elo, and Culhane (2013) have interviewed a sample of disadvantaged pregnant African-American women. They find that the median significantly underestimates the elasticity of child development with respect to maternal investments in child rearing. They report as well that remediations could have tremendous effects, since if the disadvantaged women of the sample were to correctly estimate their influence on the development of their offspring, their investments will go up 4 to 24% and the stocks of cognitive skills at age 24 months would increase between 1 and 5%.

As before, poorer parents invest less in child rearing interactions. The contribution of the dynamic extension here is to establish that income inequalities create gaps in children's accumulation of a cognitive capital. The fact that disadvantaged children are exposed to much less stimulating environments is increasingly documented and we already discussed some empirical studies on the subject in the last section. Regarding the long-term effects of

income disparities on the acquisition of cognitive skills, Rubio-Codina, Attanasio, Meghir, Varela, and Grantham-McGregor (2015) find that the significant differences that emerge in cognitive and language development among children of different socioeconomic backgrounds at age 12 months grow substantially over time. They find that the average difference of 0.53 standard deviation in cognition between the top and bottom quartiles of the wealth distribution becomes a 0.81 standard deviation difference for children aged 31-42 months. Other empirical studies find similar results, see the insightful review of Attanasio (2015).

A negative income shock negatively affects the parent's investment in child rearing and - as in the static model - may lead to a higher degree of authoritarianism along the equilibrium path. Furthermore, a negative income shock in early periods may be dramatic for the formation of a cognitive capital whenever the parent can not borrow against future incomes. Indeed, since A_t decreases as t approaches T, after a negative income shock, a parent may see a limited marginal benefit at continuing to invest in child rearing, because she knows that the child will not be able to develop enough cognitive capital anyway. In the extreme case, a negative income shock could affect so much the parent's incentive to invest in child rearing A_t that she gives up in interacting with her child. Child neglect may then follow sudden changes in economic conditions. Note that child neglect is more likely when the income of the parent is initially low, since the negative shock then makes the parent foregoes the already small benefits she had at devoting resources to child rearing.

The effects of child neglect are deleterious and long-lasting on the child's accumulation of cognitive skills. This simple prediction is attested by series of evidence showing that child neglect and deprivation are associated with delayed growth in head circumference, which directly reflects brain growth (National Scientific Council on the Developing Child (2012)). Severely neglected children also struggle when looking at human faces to correctly identify different emotions (National Scientific Council on the Developing Child (2012)). They also have significantly lower scores on language skills, higher behavior problems and have greater odds at engaging in criminal activities (Spratt, Pittenger, Swenson, Larosa, D De Bellis, Macias, P Summer, Hulsey, Runyan, and Brady (2012)). Furthermore, the later remediation is

given to deprived children, the less effective it is, as attested by the study of O'Connor, Rutter, Beckett, Keaveney, and Kreppner (2000) of adopted Romanian infants reared in orphanage.

The effect of a negative income shock on the duration of child rearing is ambiguous in the general case. As previously discussed, it can lead to child neglect, in which case the duration is drastically shorten. But if it only lowers investments without driving them to zero, then it may well delay the independency of the child, in which case it increases the overall duration of child rearing. When a negative income shock affects positively the duration of child rearing, the aggregate income that is devoted to child rearing may paradoxically increase! This is an important point because it means that besides weighting on the child's acquisition of a cognitive capital, the shock affects negatively the parent's private consumption in the long run by increasing the resources devoted to child rearing.

There is increasing evidence that early interventions aiming at improving the quality and the frequency of rearing interactions have long-lasting effects. For instance, Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) and Conti, Heckman, and Pinto (2016) show that the Perry program conducted in the mid-1960s in the district of the Perry elementary school in Ypsilanti, Michigan and the Carolina Abecedarian Project (ABC) conducted about a decade later at the Frank Porter Graham Child Development Institute significantly enhanced adult outcomes including education, employment, earnings, marriage, and participation in healthy behaviors and reduced participation in crime. Similarly, Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014) report substantial effects on the earnings of participants twenty years after a randomized intervention conducted in 1986-1987 that gave psychosocial stimulation to growth-stunted Jamaican toddlers (see as well Walker, Wachs, Grantham-McGregor, Black, Nelson, Huffman, Baker-Henningham, Chang, Hamadani, Lozoff, Gardner, Powell, Rahman, and Richter (2011) for the Jamaican data and Campbell and Ramey (1994) and Campbell, Conti, Heckman, Moon, Pinto, Pungello, and Pan (2014) for the Carolina Abecedarian Project). If little is known about the mechanisms that produce these long-lasting effects, Heckman, Pinto, and Savelyev (2013) have found that the long-run effects of the Perry school program are not essentially consequent of higher IQ levels. Indeed,

they have found that there was no statistical difference in IQ levels between male participants and non-participants and a small positive effect for female participants in a follow-up study of the Perry Preschool Program (PPP). Alternatively, a sizable portion of the adult treatment effect is explained by differences in personality traits. Indeed, treated children developed after the program norms that led to less antisocial, aggressive and rule-breaking behaviors. They also developed higher academic motivation (especially the treated girls). The fact that the PPP essentially contributed to the emergence of norms of "good" conduct is consistent with the theoretical foundations set previously on the cultural nature of the capital accumulated through rearing interactions.

1.4 Extensions

1.4.1 Child maltreatment and toxic stress as impediments to children cognitive development

The environments children grow up in shape how the brain develops. One powerful influence on brain development that we can not approach with the theory set in the previous section is children toxic stress. Toxic stress refers to events that produce strong, frequent, or prolonged activation of the body's stress management system that can physically damage brain architecture and lead to the underdevelopment of certain parts of the brain that are necessary for emotional control, memory and learning and problem-solving (National Scientific Council on the Developing Child (2005a), Shonkoff, Garner, Siegel, Dobbins, Earls, Garner, McGuinn, Pascoe, and Wood (2011)). A poor response to stress in early childhood has long lasting consequences since it affects brain development (Shonkoff, Garner, Siegel, Dobbins, Earls, Garner, McGuinn, Pascoe, and Wood (2011)). In their report on the subject, Shonkoff, Garner, Siegel, Dobbins, Earls, Garner, McGuinn, Pascoe, and Wood (2011) write "many adult diseases should be viewed as developmental disorders that begin early in life and that persistent health disparities associated with poverty, discrimination, or maltreatment could

be reduced by the alleviation of toxic stress in childhood."

A healthy nurturing environment can reduce the amount of stress children face in their early years as well as help them to cope with it. The quality of the interactions between children and caregivers is then critical in buffering children against toxic stress (National Scientific Council on the Developing Child (2004)). For instance, children whose relationships are insecure or disorganized demonstrate higher stress hormone levels and the presence of sensitive and responsive caregiver can prevent elevations in cortisol among toddlers, even in children who tend to be temporamentally fearful or anxious (National Scientific Council on the Developing Child (2005a)).

Child maltreatment and other forms of toxic stress such as domestic violence or disasters negatively affect brain development. One typical emotional functioning induced by maltreatment is the Persistent Fear Response (National Scientific Council on the Developing Child (2005a)). Children with a persistent fear response may lose their ability to differentiate between danger and safety, and they may identify a threat in a nonthreatening situation. In turn, this may be the foundation of future anxiety disorders such as PTSD (National Scientific Council on the Developing Child (2005b)). Other behavior and emotional malfunctioning associated with child maltreatment and early stress include hyperarousal, increased internalizing symptoms, diminished executive functioning, delayed developmental milestones, weakened response to positive feedback and complicated social interactions (National Scientific Council on the Developing Child (2005b)).

In order to see how maltreatment and toxic stress affect the development of cognitive abilities, we extend the model of the previous section and assume that the parent chooses three actions during the rearing interaction. She still chooses the pecuniary reward δ_t and the level of permissiveness ϵ_t, but can as well impose a cost p_t on her child when he does not behave as expected. This cost models the corporal punishments and the other maltreatments that the parent can impose on her child. We assume that it must be that p_t is such that the

utility of the child is above a survival level. The utility of the child writes

$$v^c(e_t) = \delta_t \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t u^2) du - c(e_t) - p_t(1 - \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t u^2) du) + u(e_t, S_{t-1}), \tag{1.26}$$

where S_{t-1} still refers to the cognitive capital accumulated at the beginning of period t. The maltreatment has a "rearing purpose", since it is imposed on the child when he does not behave as it is expected of him. We posit that the higher is p_t the higher is the production of toxic stress of the child in period t. Since the production of toxic stress modifies the brain chemical activity in early childhood and affects negatively the accumulation of cognitive skills, we assume that

$$S_{t-1} = \sum_{i=0}^{t-1} e_i - p_i, \tag{1.27}$$

meaning that maltreatment depreciates the accumulation of cognitive capital (with a marginal depreciation equal to 1 for simplicity). In period t, the child's recognition effort solves

$$-e_t + \frac{(\delta_t + p_t)\epsilon_t f'(e_t)}{\sqrt{2\pi(s(e_t)}} \exp(-\frac{1}{2} s(e_t)\epsilon_t^2) \leq 0, \tag{1.28}$$

meaning that it depends positively on $\delta_t + p_t$ and $\frac{\partial e_t}{\partial \delta_t} \equiv \frac{\partial e_t}{\partial p_t}$.

Assume for simplicity that the parent has an income y per period and faces the budget constraint $y \geq c_t + \delta_t$ each period. The first-order conditions associated with the optimization of the parent with respect to δ_t, p_t and ϵ_t are respectively

$$\begin{cases} \frac{\partial W(0)}{\partial \delta_t} = \alpha \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t u^2) du + \frac{\partial e_t}{\partial \delta_t}\{-\lambda_p z'(e_t) + A_t\} - u'(y - \delta_t) \\ \frac{\partial W(0)}{\partial p_t} = -\alpha(1 - \int_{-\epsilon_t}^{\epsilon_t} \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t u^2) du) + \frac{\partial e_t}{\partial \delta_t}\{-\lambda_p z'(e_t) + A_t\} - A_t \\ \frac{\partial W(0)}{\partial \epsilon_t} = 2\alpha\epsilon_t \sqrt{\frac{s_t}{2\pi}} \exp(-\frac{1}{2} s_t \epsilon_t^2) + \frac{\partial e_t}{\partial \epsilon_t}\{-\lambda_p z'(e_t) + A_t\}, \end{cases} \tag{1.29}$$

with

$$A_t = \sum_{\tau=t+1}^{T} \beta^{\tau-t}\{\alpha \frac{\partial u(e_\tau, S_{\tau-1})}{\partial S_{\tau-1}} - \lambda_p z'(e_\tau)\}. \tag{1.30}$$

Thus, from (3.24), we deduce that the parent will set a positive pecuniary reward in period t

and refrain from maltreating her child as long as

$$u'(y - \delta_t) < A_t + \alpha. \tag{1.31}$$

Indeed, if the parent's degree of altruism α is sufficiently high, *ceteris paribus*, then the marginal cost at maltreating her child is high because she suffers as well from reducing the utility of her child by maltreating him. Alternatively, if the parent's marginal benefit at increasing her private consumption is high, then she may be willing to substitute the child's pecuniary benefits for maltreatment in order to increase her own consumption. Finally, maltreatment - by increasing the child's level of toxic stress - decreases his ability to accumulate a stock of cognitive capital. Thus, a parent with a high enough marginal utility at seeing her child accumulate cognitive capital A_t will refrain from maltreating him.

According to the previous formal development then, we should expect poorer parents to rely more on maltreatment for rearing purposes. This prediction accords with the theoretical model of Weinberg (2001). Furthermore, using data from the Child Development Supplement of the Panel Study of Income Dynamics, the preceding author finds that the probability that parents use corporal punishments decline with family income and that the effect is concentrated at low income levels. Our results are consistent as well with the evidence of Straus and Stewart (1999), who show that the use of corporal punishments is more prevalent when the socioeconomic status of the parents is low.

Observe that a parent that uses corporal punishments early is more likely to keep on relying on those practices in the long run because the returns for the formation of a cognitive capital A_t fade away. Thus, we should expect maltreatment to be persistent throughout childhood. In the United States, the proportion of children who experience a report to Child Protective Services and that are re-reported is approximately equal to 25% according to the U.S. Department of Health and Human Services Administration on Children Youth and Families (2010). Our result also accords with the evidence of Proctor, Aarons, Dubowitz, English, Lewis, Thompson, Hussey, Litrownik, and Roesch (2012) that maltreated children

between ages 4 and 12 are highly likely to be abused or neglected in the future, absent intervention. Furthermore, young child age has emerged as a key risk factor. Compared to other age groups, children under the age of five have the highest rates of maltreatment reports (U.S. Department of Health and Human Services Administration on Children Youth and Families (2010)) and are at the highest risk for re-reports (Bae, Solomon, and Gelles (2009) and Fluke (2008)).

Finally, we should expect child maltreatment to correlate with lower time preferences, since then the parent does not foresee the negative effect of the toxic stress induced by such a practice on the accumulation of a stock of cognitive capital. There is evidence of a negative linkage between corporal punishments and parents' human capital, at least in recent years (see Doepke and Zilibotti (2014) for example). Again, what may be critical to explain parents' reliance on corporal punishments is not human capital per se, but the higher time preferences that correlate with it.

On a policy perspective, this extension shows how critical early interventions are, especially in social environments where children suffer from deprivation and maltreatment. From the theoretical foundations set so far and series of evidence in the related literature, it is clear that the "first-best" interventions seeking to spur children's cognitive development should not neglect early children-caregivers interactions. However, it is encouraging to observe that "second-best" interventions that simply increase parents' resources can still dramatically affect their rearing strategies by lowering their marginal utility of private consumption and thus by decreasing the likelihood of maltreatment.

1.4.2 The persistence of television viewing of young children

We have discussed in the second section of this essay how television viewing affects parents' rearing strategies and children's investments in recognition. In this section, we extend this work and study how gaps in cognitive skills can emerge from television viewing. Indeed, existing evidence points toward a negative effect of early exposure to television and video materials on cognitive outcomes. Furthermore, children that have been exposed persistently

to television throughout their early childhood are more likely to develop lower cognitive skills (Zimmerman and Christakis (2005)).

Let v_t be the time of television watching in the rearing period t (i.e. when trait w_t is taught through a rearing interaction). Assume that the appreciation of the behavioral norm at time t is

$$M_t = u(e_t, S_{t-1}), \tag{1.32}$$

with $u(.,.)$ a concave function that satisfies the assumption in (3.13), with

$$S_{t-1} = \sum_{i=0}^{t-1} z_{i,t} e_i - x_{i,t} v_i, \tag{1.33}$$

where $z_{i,t} \geq 0$ and $x_{i,t} \geq 0$. Thus, a higher time of television watching in period i depreciates the cognitive capital in norm A at time t with a marginal depreciation $x_{i,t}$. By rewriting the optimization program of the child, it is direct that his effort in period t is decreasing with the duration of television viewing in any period period τ, for $\tau \leq t-1$. Thus, there is a dynamic substitutability between previous episodes of television viewing and current investments in recognition, while there still is a dynamic complementarity between early and later investments in recognition. This difference implies that rearing strategies may be significantly more elastic to the parent's characteristics. Indeed, a parent with a low discount rate β for example will neither foresee the dynamic substitutability nor the complementarity. Thus, she may let her child watch the television and this will lead to low recognition efforts throughout the rearing period and a limited accumulation of cognitive capital. Thus, gaps in the accumulation of cognitive capital may open between the children that have been persistently exposed to television viewing and the others, because the former could not develop enough cognitive capital on standard "norms" of behavior. The linkage between television viewing and children's hyperactivity is documented for example by Ansari and Crosnoe (2016). Similarly, Christakis, Zimmerman, DiGiuseppe, and McCarty (2004) show with data from the National longitudinal Survey of Youth on children aged 1 and 3 years that hours of television viewed per day at both ages 1 and 3 was associated with attentional problems at age 7. Finally, Manganello

and Taylor (2009) show that direct child television exposure at age 1 and 3 is significantly associated with aggressive behaviors.

From the preceding theoretical foundations, we should expect poorer parents to rely relatively more on children's television viewing for rearing purposes, so as to substitute for their low investments in child rearing. By doing so however, poorer parents prevent the accumulation of cognitive capital. Interestingly then, a poor parent may persistently use the television for rearing purposes because it keeps on contributing significantly to the child's present utility given that he does not accumulate a cognitive capital. [12] Thus, inequalities in the acquisition of cognitive skills should increase, as a result of the emergence of rearing practices that heavily rely on the use of video materials and television.

1.5 Conclusion

The economic literature has been concerned in recent years with the synergies between personality traits - or character skills - and cognitive skills, as well as with the ways they are acquired and affect life outcomes.

We presented in this paper a theory of child rearing and cognitive development that starts with the premise that intergenerational rearing interactions aim at transmitting behavioral norms. Following the steps of Gary S. Becker and co-authors then, we suggested that the complementarities in the acquisition of cognitive skills documented in the literature relate to the formation of an appreciation capital for behavioral norms.

We were able to show that efficient parental rearing practices are neither permissive nor authoritarian but authoritative. Parents must express love and build a caring environment for the development of their child, while being able to condition their affection on the "good" behavior of their offspring(s). We have demonstrated that among the most important characteristics of the parent, their time preferences are central in explaining their ability to set efficient rearing strategies over the course of the rearing period, since having a higher

12. Note that an habituation to television emerges without assuming that the child builds an appreciation capital for it.

time horizon permits to exploit the dynamic complementarities inherent to the acquisition of cognitive skills and limits permissiveness, authoritarianism and child maltreatment. We also demonstrated that gaps in the acquisition of cognitive skills form between income groups because poorer parents are more authoritarian. Going further on unhealthy rearing practices, we suggested that video materials and television viewing for rearing purposes can widen the gaps in the acquisition of cognitive skills across income groups. Finally, we established a relatively simple condition that predicts the occurrence and the persistence of child maltreatment.

One key influence on the nexus between child rearing practices, cognitive development and the emergence of behavioral norms has been left aside in this paper, the influence of the prevailing social order. Indeed, rearing practices and cognitive representations are significantly context-dependent and can change rapidly, as it has been shown for instance in the studies of Patricia Greenfield and co-authors, e.g. Childs and Greenfield (1980); Greenfield, Maynard, and Childs (2003); Greenfield (2009) on the effect of globalization on cultural learning practices. Furthermore, the working of a given social group should carefully be assessed when designing early intervention programs. The framework sets in this paper could be a starting point for such future studies.

1.6 Appendix

1.6.1 Proof of Proposition 1

The first-order condition writes

$$-e + \frac{\delta f'(e)}{2\sqrt{2\pi s}} I - \frac{\delta f'(e)\sqrt{s}}{2\sqrt{2\pi}} \int_{-\epsilon}^{\epsilon} u^2 \exp(-\frac{1}{2} s u^2) du = 0, \tag{1.34}$$

with

$$I = \int_{-\epsilon}^{\epsilon} \exp(-\frac{1}{2} s u^2) du. \tag{1.35}$$

By integrating by parts the last term above, we find that

$$-e + \frac{\delta f'(e)}{2\sqrt{2\pi s}} I + \frac{\delta \epsilon f'(e)}{\sqrt{2\pi s}} \exp(-1/2 s\epsilon^2) - \frac{\delta f'(e)}{2\sqrt{2\pi s}} I = 0, \tag{1.36}$$

so

$$-e + \frac{\delta \epsilon f'(e)}{\sqrt{2\pi s}} \exp(-\frac{1}{2} s\epsilon^2) = 0. \tag{1.37}$$

The second-order condition writes

$$Z = -1 - \frac{\delta \epsilon}{\sqrt{2\pi s}} \exp(-\frac{1}{2} s\epsilon^2)[\frac{1}{2}(\epsilon^2 + \frac{1}{s}) f'(e)^2 - f''(e)] < 0, \tag{1.38}$$

which establishes the uniqueness result.

By differentiating the first-order condition, it is clear that the optimal effort is increasing with the level of reward δ. As for the effect of ϵ, we find that

$$Z \frac{\partial e}{\partial \epsilon} + \frac{\delta f'(e)}{\sqrt{2\pi s}} \exp(-\frac{1}{2} s\epsilon^2)[1 - s\epsilon^2] = 0, \tag{1.39}$$

so $\partial e / \partial \epsilon > 0$ iif $1/\epsilon > s^2$. Since $s \equiv s_0 + f(e)$, for ϵ sufficiently low, the condition $1/\epsilon > s^2$ is fulfilled and e increases with ϵ until $1/\epsilon = s^2$. Then, e decreases with ϵ and $\epsilon \to s^2$ never crosses again the curve $1/\epsilon$ because if it does, then it would mean that $e(\epsilon)$ is decreasing with ϵ for a positive measure of parameters and $1/\epsilon > s^2$, which is a contradiction.

Thus, there exists a unique threshold level of permissiveness $\overline{\epsilon}(\delta)$ such that when $\epsilon < \overline{\epsilon}(\delta)$, more permissiveness increases the child's recognition effort (i.e. e increases with ϵ). Alternatively, too much permissiveness trumps out the child's recognition effort (i.e. when $\epsilon > \overline{\epsilon}(\delta)$, e decreases with ϵ).

$\overline{\epsilon}(\delta)$ is decreasing with δ since an increase in δ shifts upward the curve $\epsilon \to s^2$, while it does not affect the curve $\epsilon \to 1/\epsilon$.

1.6.2 Proof of Proposition 2

The first-order condition with respect to δ writes

$$\frac{\partial \mathbb{E}_{a^{obs}} W}{\partial \delta} = 0 = \alpha \int_{\bar{a}-\epsilon}^{\bar{a}+\epsilon} g(u,e)du + \frac{\partial e}{\partial \delta} z'(e) - u'(y-\delta) \tag{1.40}$$

when the solution is interior. The derivative of the first term in the RHS above writes

$$2\alpha\sqrt{\frac{s}{2\pi}}\exp(-\frac{1}{2}s\epsilon^2) + \alpha\frac{\partial e}{\partial \epsilon}f'(e)\frac{\epsilon}{\sqrt{2\pi s}}\exp(-\frac{1}{2}s\epsilon^2), \tag{1.41}$$

which using (3.5) rewrites

$$\frac{\alpha e}{\delta}\{\frac{2s}{\epsilon f'(e)} + \frac{\partial e}{\partial \epsilon}\}. \tag{1.42}$$

By injecting the second-order condition (3.36) in (3.37), we can deduce that

$$\frac{\partial e}{\partial \epsilon} > \frac{2s(1-s\epsilon^2)}{f'(e)\epsilon(1+s\epsilon^2)}, \tag{1.43}$$

so

$$\frac{\alpha e}{\delta}\{\frac{2s}{\epsilon f'(e)} + \frac{\partial e}{\partial \epsilon}\} > \frac{2\alpha e s}{\delta \epsilon}\{1 + \frac{1-s\epsilon^2}{1+s\epsilon^2}\} > 0. \tag{1.44}$$

Chapter 2

A Theory on the Evolution of Religious Norms and Economic Prohibition[1]

1. I am grateful to the editor, Nathan Nunn, and to two anonymous referees for comments and suggestions that substantially improved the paper. I am particularly indebted to Thierry Verdier for helpful comments, conversations and suggestions. I thank Gani Aldashev, Alberto Bisin, Roland Bénabou, Emeline Bézin, Rachel Kranton, Sultan Mehmood and Ragnar Torvik for helpful feedback and comments. Financial support from the European Research Council under the project Tectacom 324004 is gratefully acknowledged. All mistakes are my own.

Abstract

This paper provides a new rationale for religious prohibition against secular sciences or usury. I suggest that prohibition is a strategic doctrinal innovation for religious leaders seeking to increase the diffusion of religious preferences, because it creates interferences between occupational decisions and parental investments in intergenerational cultural transmission. This theory also sheds new light on the determinants of prohibition, of its duration and of collusion equilibria between religious and political elites.

JEL codes: C73, D63, F63, Z12.

Keywords: Cultural Evolution, Cultural leaders, Religion, Political Economy, Prohibition.

2.1 Introduction

Religious cultures have persistently affected individual economic decisions and thus have undoubtedly shaped the development trajectories of our societies. The economic literature emphasizes the effect of religions on social norms that are key for economic development such as trust and cooperation, preference for education or for labor effort.[2] Nevertheless, since religions affect a plethora of beliefs - and can adapt their doctrinal views to economic changes (Carvalho and Koyama (2016); Bénabou, Ticchi, and Vindigni (2015)) - the evidence on their broad effect on growth is mixed.[3] However, a salient common feature of the major religious denominations is the repeated restrictions they have imposed upon some economic activities throughout history.[4] Both Muslim and Christian faiths have imposed bans on usury activities for example. Scientific and innovative activities encountered restrictions as well and technology adoptions were repeatedly delayed throughout history. One famous example is that of the printing press that was forbidden in the Ottoman empire until the seventeenth century. More broadly, starting with the Sunni Revival in the eleventh century, Muslim religious authorities became increasingly suspicious of secular sciences and the production of secular knowledge declined significantly (Chaney (2016)).

This paper presents a theoretical framework that aims to explain why at some point in their evolution, religions become hostile to secular sciences or usury. To this end, I develop a model of cultural evolution in a productive economy with endogenous innovation.

In the model, the agents allocate their labor in one of two productive sectors and there exist monopoly producers of sector-specific technologies. One sector is potentially subject to prohibition because it aggregates the occupations that are complementary to scientific discoveries or usury. In the former case, the monopoly producers of technologies in the

2. On the effect of religion on trust and cooperation, see Guiso, Sapienza, and Zingales (2003) and Levy and Razin (2012), on preference for education, see Becker and Woessmann (2009), Botticini and Eckstein (2012) and on labor effort, see Bénabou and Tirole (2006a) and Esteban, Levy, and Mayoral (2014).

3. Some works point to a negative effect of religiosity and a positive one for religious beliefs, see Campante and Yanagizawa-Drott (2013) and Barro and McCleary (2003) for instance.

4. Restrictions are a defining feature of religious cults, since they spur the participation in religious clubs according to the seminal work of Iannaccone (1992). See as well Berman (2000) and Aimone, Iannaccone, Makowsky, and Rubin (2013) for some evidence.

discriminated sector are scientists and innovators. For simplicity, I posit that the demand matches the supply of innovation in both sectors, which is equivalent at assuming that the production of scientific ideas perfectly diffuses into practical knowledge by creating new occupations, e.g. factory and city employments. The sector that is never subject to prohibition can be thought of as encompassing traditional occupations within the religious communities.

The model of production with endogenous innovation is embedded in a cultural evolution framework. I adopt a functionalist approach by assuming that religious norms are defined by agents' utility from belonging to communities providing social services and public goods. There is no inherent conflict between religion and secular sciences (see Section 2.2.1). Furthermore, religious norms are subject to both a decentralized evolution process where parents invest in transmitting their norms intergenerationally, and a centralized evolution mechanism where a religious leader decides whether or not to implement economic prohibition.

Finally, I build an extension of the model that accounts for the existence of collusion between religious and political elites. Doing so allows for the study of the linkages between the endorsement of political elites by religious leaders and the evolution of religious preferences, which is arguably a central concern of the latter.

The first key prediction is that during transitory periods in their evolution, economic prohibition allows religious norms to diffuse in the population because it creates a *cultural* division of labor. Indeed, a culturally segmented labor market affects the dynamics of cultural norms because it differentiates cultural groups, and consequently adds a salient dimension to parents' existing incentives to transmit their cultural norms intergenerationally. Under some conditions that I establish in the main text, adding such an economic dimension to parents' socialization decisions leads to the diffusion of higher religious preferences.

This model also has implications on the determinants of collusion between religion and politics. I find that the conditions that favor economic prohibition also foster the existence of collusion equilibria between politics and religion. Moreover, the model presented below shows that religious leaders can transfer more rents to political elites when they implement economic prohibition. Thus, the intricate linkage between cultural evolution and occupational

decisions may explain both why religious authorities become hostile to secular sciences and why religious conservatism can gain political leadership.

2.1.1 Related Literature

Economists have only recently started to study the linkage between religion and scientific production. Bénabou, Ticchi, and Vindigni (2015) put forth the "belief-eroding" effect of secular sciences and innovation on religious faiths. They show that scientific innovation can lead to three outcomes, a "Secularization" regime with unimpeded scientific progress and declining religiosity, a "Theocratic" regime with high religiosity and restrictions and an "American" regime that combines unimpeded scientific progress with stable religiosity while the church seeks to adapt its doctrine to new discoveries. This framework yields interesting insights in explaining the religious/secular divide and the co-evolution of scientific progress and religion. Bénabou, Ticchi, and Vindigni (2015) adopt an epistemological approach to religion and assume that religion and science are at least partially inherently conflictual. In this paper, I adopt an alternative view of religion and consider a functionalist approach. Sections 2.2.1 and 2.2.2 present and discuss the two approaches. Going beyond epistemology, in this paper, religions are considered as cultures coordinating social behaviors and maintaining viable communities. This departure is motivated by the fact that the existing empirical evidence on the epistemological approach is mixed (see section 2.2.1). I discuss a unique case study of secularization in Section 2.2.2 that supports the approach adopted in this paper.

To my knowledge, there has been only one previous attempt to model cultural changes in a production economy. Doepke and Zilibotti (2014) discuss the two-way link between culture and economic growth by assuming an endogenous technical change driven by innovation where agents allocate their labor to either entrepreneurial or non-entrepreneurial activities. Occupational choices depend in turn on risk tolerance or patience, which are subject to cultural transmission. The authors show that there can be multiple balanced growth paths in the long-run and that the share of entrepreneurs is higher in faster-growing countries.

This paper also contributes to the literature on the causes of sanctions from organized

religions. In a seminal article, Iannaccone (1992) argues that prohibiting or penalizing activities that compete for club members' resources limits free-riding inside the club, thus enhancing the quality of the club production which can be welfare-improving. The theory has for example been used to explain the propagation of Christianity in the Roman Empire by Stark (1996), Ultra-Orthodox Judaism by Berman (2000) and Carvalho and Koyama (2016), charitable givings in sects by James and Sharpe (2007) and the diffusion of protestantism by Hanson and Xiang (2013) (see as well Aimone, Iannaccone, Makowsky, and Rubin (2013) for some experimental evidence). Our focus in this paper is on restrictions that affected occupational decisions. They deserve special attention since they affect economic outcomes directly. One interesting result arising from the dynamic theory presented below is that in equilibrium, only the agents with weak preferences for the services provided by the religious will choose economic activities in the discriminated sectors. We should therefore expect Islamic (resp. Christian) rules banning Muslim (resp. Christian) from lending money at interest to pave the way for religious minorities to enter such occupations. This accords with the theories on occupational decisions of Jews in the Middle Ages advanced by Baron (1952) in the case of Muslim countries and by Schwarzfuchs (1966) in the European case (see Botticini and Eckstein (2012, chapter 8) for an insightful discussion on the occupational decisions of Jewish minorities during the Middle Ages).

This paper relates to the literature on forward looking cultural leaders and cultural evolution. In a recent contribution, Verdier and Zenou (2015, 2016) develop a model where individuals get socialized to specific cultural traits through two mechanisms. The first is a decentralized evolutionary mechanism consistent with the theory of cultural evolution of Bisin and Verdier (2001). The second involves cultural leaders or cultural institutions internalizing the working of the decentralized evolution process. I build on the preceding papers by assuming that cultural leaders are forward looking and exploit their knowledge of the cultural dynamics in setting their strategy. I also consider the strategy of the cultural leader over time away from the steady state of the cultural dynamics. De La Croix and Mariani (2015) consider a model of cultural conflict where cultural leaders supply and interpret culture. They focus on steady

states, not on the dynamic transition path to such steady states. Prummer and Siedlarek (2017) seek to explain the persistent differences in cultural traits of immigrant groups with the presence of community leaders. Relatedly as well, Carvalho and Koyama (2016) suggested that Jewish communities have been able to survive into the twenty-first century because they have constructed economic and social niches. They (or their leaders) have adapted their norms so as to be relatively isolated from cultural forces that led other cultural groups to abandon their traditions. In the model presented below, cultural leaders shape the labor allocation by opportunistically implementing economic prohibition so as to affect socialization decisions over time.

Beyond the preceding theoretical works, several authors have also emphasized the role of coordinated leadership in sustaining the resilience of cultural groups. Botticini and Eckstein (2012, p. 73) argue that the Pharisees, one major Jewish group, did not participate in the revolt against the Romans, allowing them to gain significant leadership over the Jewish population. They produced the basis of contemporary forms of Judaism, which no longer centered on temple service and ritual sacrifices but instead required its members to read and study the Torah (Botticini and Eckstein (2012, p. 78)). The diffusion of such a norm gave the Jews a comparative advantage in occupations for which literacy was valued in the following centuries. This may explain the persistence of Judaism in the Diaspora (Botticini and Eckstein (2012, p. 258)). Relatedly, Chaney (2016) documents a drop in the proportion of books dedicated to scientific topics in the twelfth century Islamic world. He shows that the empirical patterns are consistent with institutional changes where traditionalist religious elites gained significant leadership.

Finally, this work contributes to the literature on collusion between politics and religion. In a recent contribution, Jean-Philippe (forthcoming) provides an insightful analysis of the relationship between religion and politics in Muslim countries.[5] He argues that the decentralized structure of Islam allowed autocrats to maintain their position by instrumentalising religion. Auriol and Platteau (2017) build theoretical foundations of the relationship between

5. See as well Platteau (2008, 2011).

autocratic politics and religion in the context of non-secularized developing countries that support the analysis of Jean-Philippe (forthcoming). Other works on the collusion between religion and politics include North and Gwin (2004), Barro and Mccleary (2005), Coşgel and Miceli (2009) and Carvalho (2013).

2.2 Motivating examples

2.2.1 Beyond the epistemological debate

The epistemological conflict literature assumes that "religion and science are inherently incompatible and that a growth in science leads to decline in religion because they are competing ways of establishing truth" Engstrom and N. Engstrom (2008, p. 91). Abstracting from the possibility of collusion between politics and religion then, if religions do not adapt to the higher truth of science, they are doomed to disappear. The preceding ideas were strongly influenced by the works of Max Weber in *The Protestant Ethic and the Spirit of Capitalism* (1904) and *Economy and Society* (1933), and have been advanced in the works of leading sociologists in the 1960s and 1970s such as Peter Berger, David Martin and Brian Wilson.[6] Summarizing, in the words of Norris and Inglehart (2004, p. 7), according to the preceding school of thoughts, "the era of the Enlightenment generated a rational view of the world based on empirical standards of proof, scientific knowledge of natural phenomena, and technological mastery of the universe. Rationalism was thought to have rendered the central claims of the Church implausible in modern societies, blowing away the vestiges of superstitious dogma in Western Europe".

If science and religion are incompatible, it should be apparent in the data. More precisely, as emphasized by Iannaccone, Stark, and Finke (1996, 1998), we should expect (i) a decline in religion as scientific progress grows and lower levels of religiosity among (ii) more educated people, (iii) scientists and (iv) in "hard sciences" relative to "soft" sciences within the

6. Berger (1967), Wilson (1966) and Martin (1978). See as well the discussion on this literature in Norris and Inglehart (2004, Chapter 1).

academic community. These predictions find mixed empirical evidence. Iannaccone, Stark, and Finke (1996, 1998) provide evidence that contradicts the four preceding predictions. Evans (2011) shows that Protestant and Catholics differ from secular Americans in their propensity to seek out scientific knowledge only on the few issues where religion and science make competing claims. By contrast, Ecklund and Scheitle (2007) find that academic scientists are much less religious than the general public (consistently with (iii)), while field-specific and interdisciplinary differences are not important predictors of religiosity (in contrast to (iv)). See as well Ecklund, Park, and Veliz (2008) for similar predictions. Relatedly, Bénabou, Ticchi, and Vindigni (2015) show that religion is associated with negative attitudes toward scientific progress at the individual level.

Other lines of research have advanced the precise opposite idea that religion influenced positively science. Merton (1970, 1938) argues that certain dominant cultural values expressed in Puritanism contributed to the rise of science. Such cultural values included profitable education, empiricism and experimentation over idle contemplation (see Engstrom and N. Engstrom (2008) for a thorough discussion on the subject). This view has found supporting evidence in the work of Becker and Woessmann (2009) for example, since they find a positive effect of Protestantism on the diffusion of human capital. [7]

To conclude, the preceding series of evidence show that the epistemological conflict rationale for explaining the co-evolution of religion and science finds mixed empirical evidence in the literature. In this paper, I propose an alternative and follow a functionalist approach.

2.2.2 The process of secularization on a Danish Island

By contrast with an epistemological definition, a functionalist approach emphasizes that religions form a system of actions involving formal rituals and symbolic ceremonies, regular celebrations and the provision of social services. Religions are not primarily about explaining the physical world, but rather about coordinating social behaviors and maintaining viable

7. See as well Cantoni, Dittmar, and Yuchtman (2016) and the review on the related literature of Becker, Pfaff, and Rubin (2016).

communities. This view is consistent with the literature on religious clubs initiated in the seminal work of Iannaccone (1992). Norris and Inglehart (2004) argue relatedly that religions strive because they provide security and make people less vulnerable to risks. Consistently with this approach, Samuelson and Swinkels (2006) show that religion and welfare state spending are substitute mechanisms that insure individuals against adverse life events.

In a functionalist interpretation, understanding the interaction between science and religion implies studying how science modifies the organizational capacity of religious communities. Our key premise in that respect - which is inspired of a unique case study of secularization - is that science affects religious communities primarily when it broadens agents' economic opportunities.

Becker (1996) studies the process of religious change in a single Scandinavian county, the rural Danish island of Mors. Starting in the 1960s, the island experienced a steady decline in religiosity while scientific education and the welfare state have grown. On the surface, this dynamics seemed to confirm the standard epistemological conflict rationale for secularization. By contrast, the author shows that the decline in religiosity followed an agricultural mechanization program undertaken by Denmark shortly after the second World War. According to Becker (1996, p 439), "mechanized tractors, harvesters, feeders, and other equipment replaced most human labor in the countryside, boosting productivity. At the same time, they devastated the social system in the villages. Beginning around 1960, agricultural employment on Mors began rapidly disappearing, driving workers and their families into the town and cities in search of work. Local shops began to close, as did the voluntary associations that formed the heart of village social life".

Another important contribution of Becker (1996) lies in his analysis of the differential impact the mechanization of agriculture on the two main religious denominations present in the island, the Inner Mission and the Free Congregation. Simply put, the Inner Mission is more centered on the community than the Free Congregation. The author then shows that the latter suffered the most from the agricultural mechanization program. Becker (1996, p 440) writes "As employment disappeared, as hops and associations closed, and as members

were forced to seek work in the factories of the unconverted, the isolation of the Mission communities became impossible to maintain." Such a differential effect could not be explained with an epistemological approach, since the religious doctrine of both the Inner Mission and the Free Congregation were relatively similar, and consequently had similar likelihoods for conflicting with scientific ideas.

Finally, in a recent empirical work, Cantoni, Dittmar, and Yuchtman (2016) have identified occupational decisions to be paramount in understanding the process of secularization after of the Reformation. Cantoni, Dittmar, and Yuchtman (2016) show that following the Reformation, the closure of monasteries led to a shift toward secular occupations from university graduates from Protestant universities and a reallocation of human capital investments away from religious sector-specific theology degrees, and toward the study of more general subjects. Those patterns are consistent with a secularization mechanism instigated by the reallocation of the upper tail human capital toward secular occupations. The interaction between labor decisions and cultural evolution is the cornerstone of the theory presented in the next section.

2.2.3 The Sunni Revival

The Revival marks a period of institutional changes in the Islamic world that started sometime during the eleventh century. An intense debate between rationalist and traditionalist Muslim scholars was raging in the eve of the Revival, which is commonly associated with the defeat of rationalism (Makdisi (1994, p. 9)). Furthermore, in a recent contribution, Chaney (2016) documents a significant drop in the proportion of books written on scientific topics in the Islamic world in the period of the Revival and tests the main potential causes. His evidence buttresses the claim that an increase in the political power of conservative religious elites caused the decline in scientific output.[8] Going further, the reasons for the increased political leadership of conservative religious elites during the Revival remain to be understood.

8. The alternative explanations of the Revival include the existence of climatic shocks, the Crusades, the popularity of mystical Islam and the actions of the Turkish tribes. Chaney (2016) finds that military and climatic shocks are insufficient to explain the decline in scientific knowledge. Furthermore, he shows that the proportion of works on mysticism increases in the twelfth century and he can not rule out the effect of the Turkish tribes due to data limitations.

The theory presented in this paper allows to shed a few lights on this issue.

It seems that the Revival does not correspond to a period of political turmoil in the Islamic world. Indeed, it could have been that during the Revival, political leaders were significantly more eager to be endorsed by religious authorities.[9] Figure 3.1 replicates a figure from Blaydes and Chaney (2013). It represents the 100-years moving average of ruler duration in the Islamic world from 700 CE to 1500 CE.[10] We see that the average duration of incumbency in the Islamic world stays remarkably stable in a period that comprises the beginning of the Revival and that spans from the middle of the tenth century to the end of the twelfth century. Furthermore, the stabilization of the average duration of incumbency during the tenth century does not match the beginning of the Revival. Finally, the average duration of incumbency decreased from the thirteenth century on, suggesting that the institutional changes that accompanied the Revival did not allow political leaders to gain legitimacy in the process.

If political legitimacy was not the driving force of the Revival, then the reasons may be found in the evolution of the strategic incentives of religious leaders to collude with political elites. In that respect, the model micro-founds the incentive of religious leaders to collude with political authorities in a theory of cultural evolution. The intricate linkage between intergenerational socialization and occupational decisions provides an interpretation grid that explains both why it became optimal at some point for religious authorities to be hostile to secular sciences and why they gained political leadership during the Revival. Furthermore, the model predicts that by implementing economic prohibition, religious elites can transfer higher rents to political elites in collusion equilibria. This result suggests that traditionalist religious leaders - by using economic prohibition - could have gained a higher political leadership than their rationalist peers during the Revival. This could explain why the former won *de facto*

9. The claim that a crisis of political legitimacy can lead to collusion between political rulers and significantly more conservative religious authorities could be linked to the decentralized structure of Islam. On a related issue, see for instance Auriol and Platteau (2017).

10. Blaydes and Chaney (2013) have compiled data on the duration of rule for the highest ranking ruler in an independent governing unit in western Europe and in the Islamic world using the data from *Islamic Dynasties* (Buckser (1996)), from *Dynasties of the World* (Morby (2002)) and from the Euratlas project (Nüssli (2011)).

their epistemological dispute with the latter.

Finally, it remains to be proven that secular sciences were not targeted during the Revival because they were inherently conflicting with religious norms. To my knowledge, there were no major discoveries in the Islamic world in the eve of the Revival that could explain a significant shift in religious leaders' attitude toward secular sciences. In fact, Islamic societies were major centers for the production of scientific knowledge for centuries before the Revival. For example, $Al-Khwārizmī$ (ca. 780 - ca. 850), a mathematician of Bagdad, used geometry for the purpose of solving problems in his *Algebra*. This book circulated widely in Western Europe, and contributed to the development of symbolic algebra (Lindberg (2008, p. 487-488)). Similarly, observatories for cooperative astronomical observations were invented in the lands of Islam during the medieval era and important advances in geometrical optics were made by Islamic intellectuals (Lindberg (2008, p. 498-499)).

Although there is no evidence that scientific discoveries before the Revival were effectively diffusing into practical knowledge so as to create new economic opportunities, it seems that wealthy individuals did not solely patronized scientists for prestige, but also for practical benefits (Sagan (1996)). In the model presented below, the turning point in the history of a religious culture where economic prohibition appears is not a period where the discriminated activities threaten the religious beliefs. Rather, it is a period where their potential for broadening economic opportunities - given the cultural composition of the population - make them particularly attractive to discriminate. In that respect, the accumulation of scientific knowledge in the Islamic world before the Revival had a significant potential for broadening economic opportunities, since it contributed, centuries after, to the development of Europe.

2.3 The Model

After introducing the parameters of the model, we will first detail in section 2.3.1 the competitive equilibrium in a production economy with endogenous innovation in a given period t. In Section 2.3.2, we study the dynamics of religious preferences. The strategy of

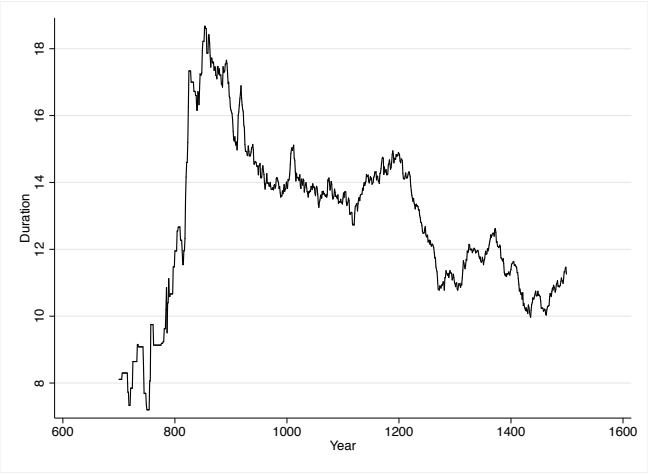

Figure 2.1 – 100-year moving average of the duration (in years) of incumbencies in the Islamic world, from Blaydes and Chaney (2013).

a religious leader is then approached in Section 2.3.3 and Section 2.3.4 studies the issue of collusion between religious and political leaders.

Consistently with a functionalist approach to religion, I posit that the religious form a community and produce collectively a non-rival but excludable good g that is positively valued by all the agents and that complements their utility from private consumption. Good g relates to the social activities undertaken by the religious group as well as to the outcome of the religious activities. I assume that there are two types of agents, the religious and the moderates. The religious and the moderates have fixed marginal utilities from consuming the religious good γ_C and γ_M respectively, with $\gamma_C > \gamma_M$ by definition. The preceding parameters are defining features of cultural types and are transmitted inter-generationally. Furthermore, I assume that γ_C and γ_M are independent from occupational decisions, meaning that there is no inherent conflict between economic activities and religious preferences.

The religious group has a unique cultural leader who (i) sets the minimum level of contribution in period t to $e_t \in [0, 1]$ that is assumed to be a linear tax on income and (ii) chooses whether or not to promote a *conservative* religious identity. By a *conservative* religious identity, I mean an identity discriminating agents on the basis of their economic occupations, say usury activities or scientific production for example. Given the strategy of their cultural leader, I assume that religious monitor the agents they meet in social interaction. If the

religious leader does not promote a conservative identity, then the monitoring only relates to individual contribution decisions. Alternatively, if the religious identity is conservative, the monitoring also relates to occupational decisions. I assume that the conservatives can exclude from the benefits of the religious good the agents that do not contribute to the religious community or that choose "bad" economic occupations. Whenever the cultural leader promotes economic prohibition, I will use interchangeably the terms conservatives and religious.

The sanction imposed upon the disrespectful agents will be referred to as social exclusion or discrimination. By contrast with the religious, the moderates are neither hostile to some economic activities nor to the agents that do not respect the basic religious requirements. They do not practice social exclusion in their daily interactions. There is a continuum $[0, 1]$ of agents. Let q_t be the fraction of religious conservatives in period t.

The labor market is divided in two sectors labeled 1 and 2 producing two goods labelled 1 and 2. The two outputs are produced with labor and technology and there is no waste of resources. For instance, one can think to a sector that encompasses the occupations that are available within the religious community, e.g. local artisans or agricultural employments while the other sector aggregates the occupations that are made available by scientific progress, e.g. factory or city employments. Furthermore, I consider that there are monopoly producers of productivity-enhancing technologies, given that technologies are inputs in the production of good 1 and of good 2. In one sector, those monopoly producers can be thought of as scientists and in the other as skilled artisans transmitting practical knowledge. I could assume instead that only one sector relies on endogenous technologies, this would not change the results.

The timing of the game is as follows. In each period t, once q_t is realized, the first decision stage is that of a forward looking religious leader. The latter has two decisions to make. He fixes the level of contribution in period t to some $e_t \in [0, 1]$ and decides whether or not to promote discriminations on the basis of economic activities. In a second stage, the agents decide simultaneously (i) whether they will respect the religious requirements and (ii) to allocate their labor in either sector 1 or sector 2. Production in both sectors occurs and the

agents consume their entire wealth (there is no inter-temporal consumption smoothing). A pairwise matching in the population follows where each religious monitors the behavior of his partner and has the ability to exclude him from the benefits of the religious good. The religious do not keep a record of past exclusions. Finally, in the last decision stage, there is a Poisson birth and death process of parameter λ and parents choose a transmission effort τ_t^i, $i \in \{C, M\}$. The same game starts at the beginning of period $t+1$ and so forth.

The time index is dropped when not necessary. There are four types of agents, since there are two economic sectors and two religious traits. I consider the following specification for the utility of an agent of cultural type $p \in \{C, M\}$ working in sector $i \in \{1, 2\}$

$$u^p(c_1^{p,i}, c_2^{p,i}, e^{p,i}) = u(c_1^{p,i}, c_2^{p,i}) v^p(e^{p,i}), \tag{2.1}$$

with $c_1^{p,i}$ (resp. $c_2^{p,i}$) his consumption of good 1 (resp. 2), $u(.,.)$ the utility from private consumption and $v^p(.)$ that from consuming the religious good g with $e^{p,i}$ the contribution to the religious good. I denote $\chi(e^{p,i})$ a parameter such that $\chi(e^{p,i}) = 1$ if the agent is discriminated by the religious and 0 otherwise, given $e^{p,i}$. Whenever the religious identity is conservative,

$$\chi(e^{p,i}) = \begin{cases} 0 & \text{if } e^{p,i} \geq e \text{ and } i = 1 \\ 1 & \text{otherwise.} \end{cases} \tag{2.2}$$

Indeed, investing an effort $e^{p,i} \geq e$ will only avoid being discriminated in sector $i = 1$, while in sector 2, $\chi(e^{p,i}) = 1$ independently from $e^{p,i}$. I assume that $v^p(.)$ is such that

$$v^p(e^{p,i}) = (1 - \chi(e^{p,i})q)\gamma_p g + u. \tag{2.3}$$

Indeed, if an agent does not contribute or work in sector 2, $\chi = 1$ so he only benefits from the religious good g with marginal utility γ_p if he does not interact with a conservative (this happens with probability q). The parameter u is assumed strictly positive and denotes the utility from consuming non-rival and non-excludable goods *not* provided by the religious.

Finally, I posit that the level of religious provision g is exogenous. In particular, I assume g independent from the aggregate level of contributions. Doing so neglects the strategy of the leader aiming at setting the level of contribution e_t, although it allows to focus on the key novelty of this model, the endogenous discrimination toward some economic activities. I assume a CES specification for $u(.,.)$,

$$u(c_1, c_2) = (\alpha c_1^{\frac{\sigma-1}{\sigma}} + (1-\alpha) c_2^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{2.4}$$

with $\alpha \in (0,1)$ a parameter which determines the taste for good 1 and $\sigma \in (1, \infty)$ the elasticity of substitution between the consumption of good 1 and good 2.

Agents only have one unit of productive time per period. I denote w_1 (resp. w_2) the wage per unit of time in sector 1 (resp. 2). Thus, the consumable wealth in a given period is $w_i(1-e)$ if a worker of sector $i \in \{1,2\}$ contributes, and w_i otherwise. The agents choose their consumption vector as well as the level of their contribution to the religious good so as to maximize their utility given in (3.1).

2.3.1 Competitive equilibrium when the religious are conservative

In order to derive the competitive equilibrium in any period t, we will proceed in three steps. First, we will derive the optimal production scheme given that output and technology producers act optimally. Next, we will study agents' optimal contributions to the religious good g at a given value of q_t, the share of conservative agents in the economy. Finally, we will combine the preceding findings in order to deduce the allocation of labor as a function of the share of conservatives in the economy.

Optimal production: The goods are produced with labor and technology, given that the levels of technology in the two sectors are endogenous to monopoly producers, e.g. artisans and scientists. I denote y (resp. $1-y$) the labor demand in sector 2 (resp. 1). The technical

level in sector i is denoted a_i so that the production frontier is given by

$$Y_i = a_i^{\beta} y_i^{1-\beta} \tag{2.5}$$

with $0 < \beta < 1$. Let $\pi(a_i)$ be the inverse demand function of technologies in sector i. The profits of a monopoly supplying sector i-specific technologies can be written as $Z_i(a_i) = (\pi_i(a_i) - c)a_i$ where c denotes a constant marginal cost (that we take equal in both sectors for simplicity). The complete derivation of the competitive equilibrium is provided in the online Appendix 1. The formalization follows Acemoglu (2002). When the firms and the agents behave optimally, we find that the relative wage in sector 2 verifies

$$\frac{w_2}{w_1} = [(\frac{1-\alpha}{\alpha})^{\sigma}(\frac{1-y}{y})]^{1/\epsilon(1-\beta)}. \tag{2.6}$$

Contribution decisions: I assume that the religious identity promoted by the leader is conservative in this section. Given the Marshallian demands for the good 1 and good 2, the indirect utility of an agent of type $p \in \{C, M\}$ working in sector $i \in \{1, 2\}$ is given by

$$v^{p,i}(e^{p,i}, p_1, p_2, w_i) = w_i(1 - e^{p,i})[(1 - \chi(e^{p,i})q)\gamma_p g + u]z(p_1, p_2), \tag{2.7}$$

with

$$z(p_1, p_2) = (\alpha^{\sigma} p_1^{1-\sigma} + (1-\alpha)^{\sigma} p_2^{1-\sigma})^{1/\sigma - 1}. \tag{2.8}$$

First, it is clear that workers in sector 2 have no incentive in contributing to the religious good, since they always face discriminations. Thus, $e^{p,2} = 0$ for $p \in \{C, M\}$.

For the sector 1 workers, contributing can be optimal whenever the cost of being discriminated is too high. Indeed, it is optimal to set $e^{p,1} = e$ whenever $v^{p,1}(e, p_1, p_2, w_1) \geq$

$v^{p,1}(0, p_1, p_2, w_1)$. From (3.11), the preceding inequality implies

$$e^{p,1} = \begin{cases} e & \text{if } (1-e)(\gamma_p g + u) \geq (1-q)\gamma_p g + u \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

The inequality in (3.13) gives a minimum fraction of conservative $q(\gamma_p)$ such that the agents of type $p \in \{C, M\}$ working in sector 1 will contribute when $q \geq q(\gamma_p)$,

$$q(\gamma_p) = e(1 + \frac{u}{\gamma_p g}).$$

Note that $q(\gamma_p)$ increases with the minimum religious requirement e. The threshold $q(\gamma_p)$ increases as well with the reservation utility u, while it decreases with the utility derived from the consumption of the religious good g. Finally, $q(\gamma_p)$ decreases with γ_p since the higher the taste for the religious good, the higher the cost of being socially excluded. In particular, $q(\gamma_C) < q(\gamma_M)$ holds since $\gamma_C > \gamma_M$, meaning that religious will contribute for a strictly larger set of parameters. I will denote $q(\gamma_p) \equiv q_p$ to simplify the notations.

Summarizing the preceding findings, if $q < q_C$, no one contributes, if $q_C \leq q \leq q_M$, while working in sector 1, conservatives do contribute while moderates do not. Finally, when $q > q_M$, any agent working in sector 1 contributes. The difference in contribution decisions in the region where $q \in (q_C, q_M)$ provides a first incentive for transmitting cultural traits inter-generationally (see for instance Levy and Razin (2012) for a similar motive of inter-generational transmission).

Occupational decisions: As a simple illustration, I first describe the equilibrium on the labor market whenever the religious do not discriminate the agents working in sector 2. Given the expression of the indirect utility (3.11), the agents face the same incentive to contribute in both economic sectors, so moderates (resp. religious) will contribute whenever $q \geq q_M$ (resp. $q \geq q_C$), independently from their occupation. I denote e^M (resp. e^C) the contribution of moderates (resp. conservatives).

Regarding the optimal allocation, moderates (resp. conservatives) are indifferent between sector 1 and sector 2 whenever $v^{M,1}(e^M, p_1, p_2, w_1) = v^{M,2}(e^M, p_1, p_2, w_2)$ (resp. $v^{C,1}(e^C, p_1, p_2, w_1) =$

$v^{C,2}(e^C, p_1, p_2, w_2)$). From (3.11), both types are indifferent whenever the wages are equal in the two sectors, $w_2 = w_1$. Thus, substitution $w_2/w_1 = 1$ in (2.6) gives the equilibrium allocation of labor y_0. I find that

$$y_0 = \frac{(1-\alpha)^\sigma}{(1-\alpha)^\sigma + \alpha^\sigma},$$

meaning that the size of sector 2 reflects the taste of the consumer for the output produced in that sector. The preceding textbook results will not hold whenever conservative religious discriminate sector 2 workers, since this will instigate a wage premium in sector 2 that compensates for the loss implied by social exclusion. This will negatively affects the demand for labor in sector 2. Moreover, such a distortion will create conflicting views between moderates and conservatives over the optimal allocation of labor, resulting in sorting effects on the labor market that will ultimately affect inter-generational transmission decisions.

Let $y_M(q)$ (resp. $y_C(q)$) be the size of sector 2 that makes moderates (resp. conservatives) indifferent between working in the two sectors whenever the fraction of conservatives in period t is equal to q.

Lemma 1. *For $p \in \{C, M\}$, the allocation $y_p(q)$ is piecewise continuous over the segment $[0, 1]$,*

- *if $q \leq q_p$, the agents of type p are indifferent between the two sectors for a wage premium in sector 2 $\eta_p(q) = w_2/w_1 = 1$ and $y_p(q) = y_0$ necessarily.*
- *if $q > q_p$, the agents of type p are indifferent between the two sectors for a wage premium in sector 2*

$$\eta_p(q) = \frac{(1-e)(\gamma_i g + u)}{(1-q)\gamma_i g + u}, \tag{2.10}$$

with $\eta_C(q) > \eta_M(q) > 1$. The allocation $y_p(q)$ is non-increasing with q,

$$y_p(q) = \frac{(1-\alpha)^\sigma}{(1-\alpha)^\sigma + \alpha^\sigma \eta_p(q)^{\epsilon(1-\beta)}}. \tag{2.11}$$

- *$y_p(q)$ decreases with g and γ_p and increases with e and u. $y_C(q) \leq y_M(q)$ with a strict*

inequality for $q > q_C$.

If $q \leq q_p$, then type p workers derive the same utility from consuming the religious good in the two sectors and they occasionally suffer social exclusion. Consequently, type p workers will find it optimal in equilibrium that $w_1 = w_2$, i.e. no wage premium in sector 2, which from (2.6) gives $y_p(q) = y_0$.

If $q > q_p$, workers of type p will be indifferent between the two sectors whenever the wage premium in sector 2 equilibrates (i) the loss implied by social exclusion in sector 2 with (ii) the linear income tax implied by paying the contribution to the religious while working in sector 1. That is to say, if type p is indifferent between the two sectors, $v^{p,1}(e, p_1, p_2, w_1) = v^{p,2}(0, p_1, p_2, w_2)$ necessarily holds, and we deduce with (3.11) that $w_2/w_1 = \eta_p(q)$. Since contribution decisions are made optimally (one contributes only when the cost of being discriminated is too high, or equivalently $q > q_p$), then $\eta_p(q) > 1$ necessarily holds. Since the cost of being socially excluded is higher for conservatives (given that they have a higher taste for the religious good), they will need a higher wage premium in sector 2 in order to be indifferent between the two sectors, $\eta_C(q) > \eta_M(q)$. We deduce $y_p(q)$ by substituting $w_2/w_1 = \eta_p(q)$ in (2.6).

The loss implied by social exclusion increases with q and so does $\eta_p(q)$. Intuitively, workers of type p will prefer higher wages in sector 2 to compensate for a higher cost of being excluded when q is high. Since the inverse demand functions for labor are downward sloping (see (2.6)), then the equilibrium fraction of the labor force employed in sector 2 $y_p(q)$ decreases with q.

The last point of the proposition follows directly from the variations of the wage premium $\eta_p(q)$ with the parameters. If g increases, then it becomes more costly to be socially excluded, so the agents want higher wages in sector 2. This affects negatively the demand for labor in sector 2, which results in a lower equilibrium fraction of the labor force in that sector $y_p(q)$. The reasoning is exactly the same for an increase of γ_p, and the opposite for an increase of e or u. Finally, observe that conservatives derive a higher utility in sector 1 relative to moderates when both types pay the religious contribution e because $\gamma_C > \gamma_M$. This implies that conservatives require a higher wage premium in sector 2, $\eta_C(q) > \eta_M(q)$. Since the inverse

demand functions for labor are downward sloping, this also means that if conservatives are indifferent between the two sectors, then the fraction of the labor force in sector 2 is lower, $y_C(q) < y_M(q)$.

Before turning to the study of the sorting effects on the labor market, it is worth detailing the effect of social exclusion of sector 2 workers on the relative technical progress. Since labor and technology are complementary in the production of the two outputs, the inverse demand function for technology is shifted downwardly whenever the labor demand for workers in sector 2 is (this is the market size effect of Acemoglu (2002)). When the outputs produced in the two sectors are gross substitutes (i.e. $\sigma > 1$), a lower labor force in sector 2 will bias technical progress toward sector 1 (see Acemoglu (2002) and the Online Appendix 1). Indeed, if the allocation of labor is equal to $y_p(q)$, then the relative technical level in sector 2 is such that

$$a_p(q) \equiv \frac{a_2}{a_1} = \eta_p(q)^{-(1-\beta)(\sigma-1)}\left(\frac{1-\alpha}{\alpha}\right)^{\sigma}, \tag{2.12}$$

which is strictly decreasing in the wage premium $\eta_p(q)$. Furthermore, when the wage premium is equal to 1, then the religious create no distortion on the direction of technical change. Thus, whenever $q < q_C$, it follows from Lemma 2 that the conservatives are not sufficiently numerous to create labor distortions, so they have no effect on the production of technologies.

Proposition 1.

(i) Labor allocation: The equilibrium demand for labor in sector 2 $y(.)$ is a piecewise continuous function of the fraction of conservative q, $y(q) = \max(\min(1-q, y_M(q)), y_C(q))$ if $q > q_C$ and $y(q) = y_0$ otherwise.

(ii) Sorting effect: Whenever $y(q) = y_M(q)$ (resp. $y(q) = y_C(q)$), moderates (resp. conservatives) populate both sectors while conservatives (resp. moderates) are present only in sector 1 (resp. 2). Whenever $y(q) = 1 - q$, there is a perfect sorting on the labor market, conservatives (resp. moderates) populate sector 1 (resp. 2).

In order to grasp the intuition, note that the characterization of $y(q)$ in the first point of

Proposition 2 can be rewritten as

$$y(q) = \begin{cases} y_M(q) & \text{if } y_C(q) \leq y_M(q_t) \leq 1-q \\ 1-q & \text{if } y_C(q) \leq 1-q \leq y_M(q) \\ y_C(q) & \text{if } 1-q \leq y_C(q) \leq y_M(q) \end{cases} \tag{2.13}$$

when $q > q_C$. First, in the case where $y_C(q) \leq y_M(q_t) \leq 1-q$, moderates are not constrained by the size of their group and thus will migrate in sector 2 until being indifferent, so $y(q) = y_M(q)$. Regarding the sorting effect, whenever $y(q) = y_M(q) < y_0$, note that it is strictly more beneficial for conservatives to be in sector 1 than in sector 2, so they all migrate in the former.

Second, when $y_C(q) \leq 1-q \leq y_M(q)$, moderates are constrained by the size of their group and thus $y(q) = 1-q$. Interestingly in this case, as occupational choices are not made cooperatively, there will be no conservative in sector 2. The intuition is that moderates will all migrate in sector 2 because $1-q \leq y_M(q)$. Given that $y_C(q) \leq 1-q$, the best-response of conservatives is not to enter sector 2. By doing so, they keep the size of that sector as close to y_C as possible. Social exclusion implies a perfect sorting on the labor market in that case.

When $1-q \leq y_C(q) \leq y_M(q)$, moderates are so constrained by the size of their group that conservatives can reach their optimal allocation $y(q) = y_C(q)$. In this case, all the moderates will be in sector 2 and conservatives will occupy both sectors. This is because moderates have a strictly larger utility in sector 2 given that the share of worker in that sector is strictly lower than their indifference point at $y_M(q)$. The labor allocation $y(.)$ as a function of q is depicted in the right panel of figure 3.2.

2.3.2 The dynamics of religious preferences

We denote

$$\begin{cases} u^M(q_t) = \max(v^{M,1}(e_t^{M,1}, p_1, p_2, w_1), v^{M,2}(e_t^{M,2}, p_1, p_2, w_2)) & \text{and} \\ u^C(q_t) = \max(v^{C,1}(e_t^{C,1}, p_1, p_2, w_1), v^{C,2}(e_t^{C,2}, p_1, p_2, w_2)). \end{cases} \tag{2.14}$$

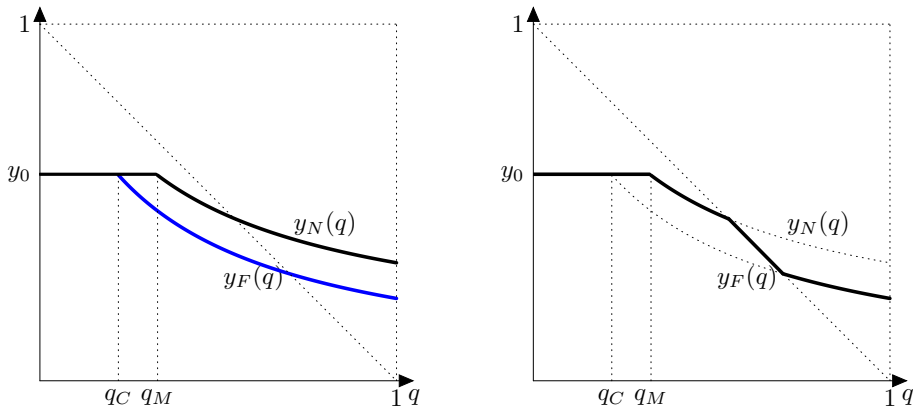

Figure 2.2 – Labor allocation as a function of the fraction of religious q.

The indirect utility functions in the preceding expressions are given in (3.11), and the prices are at their equilibrium values so they depend only on the fraction of conservatives q_t. Since $z(p_1, p_2)$ enters multiplicatively the indirect utilities - independently from cultural types and occupational decisions - it will not affect the dynamics. I assume that $z(p_1, p_2) \equiv 1$ in the sequel. $u^M(q_t)$ (resp. $u^C(q_t)$) is then the utility derived by the moderates (resp. conservatives) in period t given a general equilibrium in production and consumption in that period. Indeed, if the moderates populate both sectors in period t, then $y(q_t) = y_M(q_t)$ and they reach the same utility in both sectors by definition so they all derive a utility $u^M(q_t)$. Alternatively, if the moderates are only in sector 2, then again they obviously all get a utility $u^M(q_t)$. The same reasoning holds for $u^C(q_t)$.

Consider an overlapping generation structure. We are concerned with the evolution of the preferences for the tastes γ_i for the religious good, $i \in \{C, M\}$. Following the steps of Bisin and Verdier (2000b, 2001) in the Online Appendix 3.6.2, we find that the dynamics of religious preferences is such that

$$\dot{q}_t = \lambda q_t(1 - q_t)(\tau^C(q_t) - \tau^M(q_t)), \tag{2.15}$$

with

$$\begin{cases} \tau^C(q_t) = (1-q_t)(u^{CC}(q_t) - u^{CM}(q_t)) \\ \tau^M(q_t) = q_t(u^{MM}(q_t) - u^{MC}(q_t)). \end{cases} \tag{2.16}$$

The function $u^{ij}(.)$ gives the utility of being a type $j \neq i$, as evaluated by a parent of type i under the imperfect empathy assumption and the function $u^{ii} \equiv u^i(.)$ is given in (3.21).

At this stage, we have to differentiate four cases along the lines of Proposition 2 in order to compute the values of the $u^{ij}(q_t)$ parameters, $i, j \in \{C, M\}$. Indeed, given the value of q_t relative to the thresholds q_C and q_M, the cultural types will adopt different contributing behaviors while working in sector 1. Moreover, given the value of $y(q_t)$, cultural types may or may not be indifferent between the two sectors and this will side effect on cultural transmission decisions. As a simple illustration, when $y(q_t) = y_M(q_t)$ and $q_t > q_M$, then $\tau^M(q_t) = 0$ necessarily (see case 2.1 in the Online Appendix 2). Indeed, since moderates are indifferent between the two sectors, there is a sorting effect where conservatives only populate sector 1 which grants strictly higher utility levels from their perspective. This makes them invest in inter-generational transmission, while moderates are indifferent so they do not invest in transmitting their cultural trait. $\dot{q}_t > 0$ necessarily in the region $q_M < q_t < \underline{q}$, as represented in figure 3.3. I call this an *economic effect* in cultural transmission decisions. [11] By the same token, observe that when $y(q_t) = y_C(q_t)$, then moderates invest in inter-generational transmission while conservatives do not. Thus, $\dot{q}_t < 0$ for $q_t > \overline{q}$ (see figure 3.3).

I assume in the rest of the paper that the taste parameter α is not disproportionally low, meaning that the fraction of the labor force that should be in sector 2 whenever there is no discrimination is not too high and the condition $y_0 < 1 - q_M$ holds. This assumption is useful because it implies that in the region $q_C < q_t < q_M$, $y(q_t) = y_M(q_t)$ necessarily holds, in direct application of Proposition 2 (this can be seen on Figure 3.2). This, in turn, simplifies the dynamic analysis since it means that in the region $q_C < q_t < q_M$ the economic motives are absent (see case 2.2 in the Online Appendix 2).

11. While working in sector 1, the agents contribute independently from their cultural type since $q_t > q_M > q_C$.

Proposition 2. *Assuming that* $y_0 < 1 - q_M$*,*

- *if* $q_t < q_C$*, there is no cultural evolution,* $\dot{q}_t = 0$*.*
- *if* $q_C < q_t < q_M$*,* $y(q_t) = y_M(q_t) = y_0$*, there exists a unique unstable rest point of the cultural dynamics* $q_U \in (q_C, q_M)$*.*
- *if* $q_t > q_M$*, whenever* $y(q_t) = y_C(q_t)$*, the fraction of conservatives necessarily decreases over time,* $\dot{q}_t < 0$*. Alternatively, whenever* $y(q_t) = y_M(q_t)$ *the fraction of conservatives necessarily increases over time,* $\dot{q}_t > 0$*. Moreover, in any convex set* $[\underline{q}, \overline{q}]$ *such that* $y(q_t) = 1 - q_t$*, there can exist several stable rest points of the cultural dynamics.*

The proof of the unicity claims in the region of parameters $q_C < q_t < q_M$ is given in the Online Appendix 2. First, observe that in the region $q_C < q_t < q_M$, the difference in contribution behaviors is the only motive of cultural transmission strategies (because there is no wage premium in sector 2). A substitution and a complementarity effects drive the cultural dynamics. The former is standard in the economic literature on cultural transmission (e.g. Bisin and Verdier (2001)). The complementarity effect stems from social exclusion. The higher q_t, the higher the cost of being excluded so the lower (resp. higher) the incentive of moderates (resp. conservatives) to transmit their trait. I show in the Online Appendix 2 that the complementarity effect dominates. As represented in figure 3.3 then, $\dot{q}_t$ increases with q_t in the region $q_C < q_t < q_M$.

The claim that there can exist several stable rest points of the cultural dynamics in any convex set $[\underline{q}, \overline{q}]$ such that $y(q_t) = 1 - q_t$ comes from the fact that the economic motive can counterbalance the complementarity effect. Indeed, the cost of social exclusion is internalized in labor decisions. It reduces the supply of labor in sector 2, and consequently instigates a wage premium in that sector that shifts upward the transmission effort of moderates. Thus, the cost of being excluded can be entirely compensated by the wage premium in sector 2. In other words, the economic motive can neutralize the complementarity effect arising from social discrimination. This is why the cultural dynamics can admit stable rest points as well. I give sufficient conditions for existence and uniqueness in the next proposition.

Proposition 3. *Whenever* $y_0 < 1 - q_M$*,*

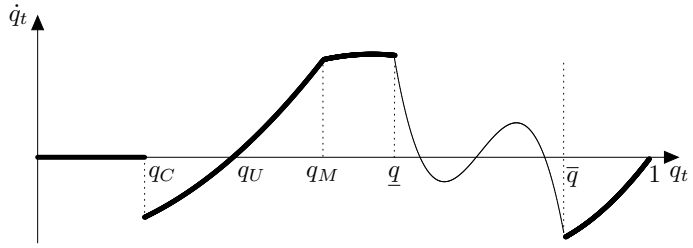

Figure 2.3 – Phase diagram.

(i) There exist two parameters $\underline{q}$ *and* $\overline{q}$ *in* $(0,1)$ *such that* $y(q_t) = y_M(q_t)$ *if* $q_t < \underline{q}$, $y(q_t) = 1 - q_t$ *if* $\underline{q} < q_t < \overline{q}$ *and* $y(q_t) = y_C(q_t)$ *if* $\overline{q} < q_t$.

(ii) The condition $y(q) = 1 - q$ *defines a unique interval* $[\underline{q}, \overline{q}]$ *where the cultural dynamics admits at least one stable rest point* $q^* \in (\underline{q}, \overline{q})$. *A sufficient condition for uniqueness is* $\sigma < \hat{\sigma}$, *with*

$$\hat{\sigma} = \frac{1}{1-\beta}\{\frac{1}{4}\frac{\gamma_C + u/g}{2\gamma_C - \gamma_M} - \beta\}.$$

The assumption $y_0 < 1 - q_M$ is sufficient for the existence of stable rest points. Indeed, the claim that when $y_0 < 1 - q_M$ there exist two parameters $\underline{q}$ and $\overline{q}$ in $(0,1)$ such that $y(q_t) = y_M(q_t)$ if $q_t < \underline{q}$, $y(q_t) = 1 - q_t$ if $\underline{q} < q_t < \overline{q}$ and $y(q_t) = y_C(q_t)$ if $\overline{q} < q_t$ is direct from the monotonicity of $y_M(.)$ and $y_C(.)$. The fact that there exists at least one stable rest point whenever $y(q) = 1 - q$ follows from a continuity argument. Indeed, at the limit where $q_t = \underline{q}$, $y(q) = y_M(q)$, so $\dot{q}_t > 0$. At the opposite, at the limit where $q_t = \overline{q}$, $y(q) = y_C(q)$, so $\dot{q}_t < 0$ necessarily. Thus, by continuity, there exists at least one stable rest point $q^* \in (\underline{q}, \overline{q})$.

As detailed in the Online Appendix 2, when $y(q_t) = 1 - q_t$, the relative effort of transmission of the conservatives $\Delta\tau^C(q_t) = (\tau^C(q_t) - \tau^M(q_t))/w_1$ writes

$$\Delta\tau^C(q_t) = (1-e)[g\{(1-q_t)\gamma_C + q_t\gamma_M\} + u] - \eta(q_t)\{u + g[(1-q_t)^2\gamma_C + q_t(1-q_t)\gamma_M]\}. \quad (2.17)$$

The first term in the RHS of (2.17) is decreasing with q_t, since it relates to the substitution effect only. The second term in the RHS of (2.17) shows the tension between (i) the combination of the economic and the substitution effects and (ii) the complementarity effect. Indeed, $\eta(q_t) = w_2/w_1 > 1$ is the wage premium in sector 2, and is strictly increasing in q_t since the

scope of discrimination increases with q_t. This economic effect tends to decrease the relative effort of transmission of the conservatives, since the latter only populate sector 1. However, the higher is q_t, the more social pressure there is on the moderates. This tends to increase $(\tau^C(q_t) - \tau^M(q_t))/w_1$. The dominant effect will in turn depend on the elasticity of the wage premium with respect to q_t, which is not surprisingly a function of the elasticity of substitution between the two consumption goods.

I establish in the Online Appendix 3 a sufficient condition for uniqueness. In particular, I find that whenever σ is sufficiently low, the complementarity effect in cultural transmission strategies is dominated by the economic effect. Intuitively, whenever σ is low, the consumers have a low willingness to substitute their consumption of good 2 for that of good 1. Consequently, even when the fraction of conservatives is high, the firms producing good 2 still supply relatively high quantities of that good despite the price of labor being higher - since it compensates for the cost of social exclusion.

2.3.3 Perfectly forward looking religious leader

We now have sufficiently developed the model to study the decision stage of the religious leader. Let $d_t \in \{0, 1\}$ the choice variable of the religious leader such that $d_t = 1$ if economic prohibition is implemented in period t and $d_t = 0$ otherwise. I assume that the leader seeks to maximize a stream of rents, which are proportional to the size of the cultural group. Let W be the rent that the religious leader obtain when there are more individuals in the population who are like him, i.e. who adopt the same values. The program of the leader can thus be written in the form:

$$\max_{\{d_t\}_{t\geq 0}} \int_0^\infty e^{-\rho t}(W q_t - c d_t)dt$$
$$\text{s.t.} \begin{cases} \dot{q}_t = \lambda q_t(1 - q_t)(\tau^C(q_t) - \tau^M(q_t)) \\ q_0 = q(0), \end{cases} \tag{2.18}$$

where $\rho > 0$ is the time preference of the leader. I assume the rent W to be exogenous from the other parameters of the model for simplicity. The parameter c corresponds to the cost per

period from implementing and maintaining economic prohibition.

In order to gain intuition on the feasible strategies for the religious leader, I represent two typical dynamics in Figure 3.4. The upper panel represents the dynamics when there is no economic prohibition. The labor allocation is necessarily y_0 independently from the cultural composition of the population, and there is no wage premium in sector 2. Consequently, there will be no economic motive in socialization decisions. In the region $q_C < q_t < q_M$, the conservatives contribute while the moderates do not. This difference induces different socialization efforts, which account for both a substitution effect (e.g. Bisin and Verdier (2000b) and a complementarity effect (see Proposition 2). It is shown in the proof of Proposition 2 that the complementarity effect dominates, meaning that the higher the size of the conservative group, the higher the incentive of conservative parents to socialize their offspring to their own trait. This leads to an increasing population of conservatives, up to the point where q_t reaches q_M. Then, once $q_t > q_M$, since both cultural groups adopt the same behavior, the parents have no incentive to invest in transmitting their cultural trait inter-generationally and the dynamics of religious preferences stop.

The lower panel gives the dynamics of religious preferences when the leader decides to implement economic prohibition. Observe that in the region where $q_M < q_0 < q^*$, while $\dot{q}_t = 0$ in the upper panel, $\dot{q}_t > 0$ in the lower panel. This is because economic prohibition creates an economic motive in cultural transmission, as described earlier. Thus, it can be particularly interesting for a cultural leader to promote economic discrimination when the initial size of the group q_0 is such that $q_M < q_0 < q^*$, because the economic motive will segment the labor allocation, and will consequently affect socialization decisions in such a way that the size of the religious group grows over time.

Proposition 4. *Whenever $q_M < 1 - y_0$ and $\sigma < \hat{\sigma}$, denoting q_0 the initial size of the conservative group,*

(i) if $q_0 < q_U$, then economic prohibition is not implemented and q_t reaches q_C in some finite period. Similarly, if $q_0 > q^*$, economic prohibition is never implemented and q_t stays at q_0 indefinitely.

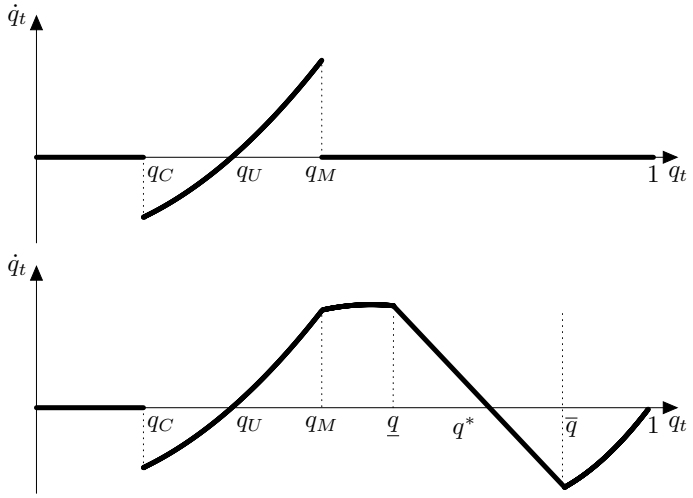

Figure 2.4 – Upper (resp. lower) panel: phase diagram when the religious leader does not (resp. does) implement economic discriminations.

(ii) if $q_U < q_0 < q^*$, economic prohibition is - if implemented - transitory. It eventually appears in some finite period $t_b(q_0)$ and then necessarily becomes obsolete in some finite period $\tilde{t}(q_0) + t_b(q_0)$. Furthermore, given that c is sufficiently low, economic prohibition necessarily appears in the history of religions and becomes obsolete when q_t reaches some value $q(\tilde{t}(q_0) + t_b(q_0)) < q^*$.

(iii) The duration of economic prohibition $\tilde{t}(q_0)$ increases with the time preference of the leader ρ, with the rents W, with the taste for the religious good γ_p, $p \in \{C, M\}$ and decreases with the cost of implementation c and with the reservation utility u.

The assumptions of the proposition insure that the steady state of the dynamics when economic prohibition is implemented is unique (see Proposition 4). The key result in Proposition 4 says that economic prohibition is necessarily transitory in the history of religions. The intuition of this result is the following. First, if a religious group has initially a low size, then the social pressure induced by economic prohibition will not be sufficient to alter labor outcomes. Thus, there will be no "economic motive" in socialization decisions that will affect the dynamics of religious preferences whether economic prohibition is or is not implemented. It is then suboptimal for the religious leader to pay the cost of discriminations c when his preferences are not sufficiently diffused in the population. But notice that the

difference in contributing behaviors between the two groups can trigger complementarities between socialization decisions and the dynamics of religious preferences that leads to the growth of the conservative trait. Indeed, when $q_U < q_0 < q_M$, the complementarity effect will make the conservative group grow over time. It is only when q_t reaches q_M that the question of implementing economic prohibition becomes salient for the religious leader.

If he decides to implement economic discrimination, he knows that the labor structure will become segmented, and that this will create an economic motive in socialization decisions that makes the fraction of conservative grow toward *at most* q^*, since at q^* the dynamics stop (see the lower panel of figure 3.4). Assume momentarily that the leader maintains economic discrimination until q_t reaches q^*. It is direct that prohibition becomes obsolete when q_t reaches q^* because it does not allow to increase the fraction of conservative (a steady state of the dynamics is reached), while it implies a strictly positive cost. Thus, mature religious groups should stop discriminating on the basis of economic occupations.

Observe furthermore that there is a trade-off between short-term costs and long-term benefits inherent to the determination of the duration of economic prohibition. On the one hand, the religious leader may prefer not to pay the cost c of implementing economic discrimination, given that the size of the group is sufficiently large. On the other hand, implementing economic discrimination might allow to reach a larger group size in future periods. In order to describe this trade-off more precisely, take the case where $q_U < q_0 < q^*$. The maximization program of the leader can be rewritten in the form

$$\max_{\tilde{t}\geq 0} W(\tilde{t}) \quad \text{s.t.} \begin{cases} \dot{q}(t) = \lambda q(t)(1-q(t))(\tau^C(q(t)) - \tau^M(q(t))) \\ q(0) = q_0, \end{cases} \tag{2.19}$$

with

$$W(\tilde{t}) = \int_0^{t_b(q(0))} e^{-\rho t} W q(t) dt + \int_{t_b(q(0))}^{t_b(q(0))+\tilde{t}} e^{-\rho t} (W q(t) - c) dt + \frac{e^{-\rho(\tilde{t}+t_b(q(0)))}}{\rho} W q(\tilde{t} + t_b(q_0)). \tag{2.20}$$

This says that the decision of the leader can be simplified to a simple determination of the duration of economic prohibition. Indeed, from period 0 to some finite period $t_b(q(0))$, $q_U < q_t < q_M$ and it is optimal not to implement economic discrimination as explained previously (the duration $t_b(q(0))$ is defined by $q(t_b(q(0))) = q_M$, see the first term in the RHS of (2.20)). Once $q(t)$ reaches q_M, then economic discrimination becomes optimal for a period $\tilde{t}$ and is obsolete whenever $q(t)$ reaches $q(t_b(q(0)) + \tilde{t})$ (see the second and the third terms in the RHS of (2.20)). The first-order condition associated to this problem simplifies to

$$e^{-\rho(\tilde{t}+t_b(q_0))} \{ \frac{W}{\rho} \dot{q}(\tilde{t} + t_b(q_0)) - c \} \leq 0, \tag{2.21}$$

which holds with equality for $\tilde{t} > 0$. In setting the duration of economic prohibition, the leader faces a simple trade-off. If he extends marginally $\tilde{t}$, he will pay a cost c in period $\tilde{t} + t_b(q_0)$. However, discrimination will guarantee that the conservative cultural group grows - given that it affects parents' socialization decisions (i.e. $\dot{q}(\tilde{t} + t_b(q_0))$). When c is sufficiently low, it is shown in the Online Appendix that there will necessarily be economic prohibition in equilibrium, because the LHS of (2.21) is strictly positive for $\tilde{t} = 0$. This is because the speed of cultural change is so high that it is always optimal for the leader to pay the cost of economic prohibition initially. However, as $q(t)$ approaches the steady state of the dynamics, then the speed of cultural change becomes excessively low. Thus, at some point, the leader will prefer to stop discriminating economic activities. Moreover, the size of the religious group will necessarily be such that $q(\tilde{t} + t_b(q_0)) < q^*$ in the long run. Simply put, it is optimal to stop economic discrimination before the steady state is reached because at this point the speed of cultural change is precisely equal to zero, so the marginal cost of prohibition is strictly above the marginal benefit.

The comparative statics are also a direct consequence of (2.21). When the cultural leader becomes more patient, i.e. ρ decreases, then he enjoys seeing the size of the group increase relatively more and will consequently discriminate economic activities for a larger time period. Similarly, any factor that affects positively the speed of cultural evolution $\dot{q}(\tilde{t} + t_b(q_0))$ will increase the duration of economic prohibition. Whenever the agents have a higher taste for the religious good for instance, then $\dot{q}(\tilde{t} + t_b(q_0))$ increases because the relative effort of cultural transmission of the conservatives does (see Proposition 4). The reasoning is similar for the rest of the comparative statics of Proposition 4.

If economic prohibition eventually appears in the history of religions, it necessarily lasts for a limited period of time, which is affected by key socio-economic parameters. In societies where religious conservatives have the ability to provide public services to the population at large for instance, we should expect higher durations of economic prohibition. This is because conservative leaders anticipate that segmenting the labor market allows their cultural trait to diffuse in the population. Similarly in highly religious societies - and given that religious leaders possess the mean to provide social services, economic prohibition should be expected. This accords with the casual evidence that economic prohibition has been widely used for centuries in the pre-industrial era when organized religions where paramount in the prevailing institutional arrangements.

Observe also, again from the lower panel of Figure 3.4, that when the group of conservatives is sufficiently large initially, then any intervention of the cultural leader will affect socialization decisions so as to decrease the size of the conservative group. This is because any difference in the economic behaviors of the two groups will incentivize the parents to invest in inter-generational transmission, but their efforts of socialization will be heavily affected by the substitution effect (e.g. Bisin and Verdier (2000b)) that favors the moderates, since the conservatives are numerous. Consequently, in order to avoid the cultural substitution, the cultural leader will prefer not to create a distortion in socialization decisions. Finally, when q_t is initially significantly low, i.e. $q_0 < q_U$, then there is no point in implementing economic discrimination because the size of the group is so small that it does not affect labor outcomes,

and anyway, the size of the conservative group shrinks over time since the cost of contributing to the religious good is larger than the expected loss from social exclusion.

From the previous paragraphs, we can deduce that economic prohibition will eventually arise if the size of the religious group is high enough so that discriminations affect the labor allocation on the first hand, but not overwhelmingly high, so that the induced sorting effects favor the spread of the religious trait on the other hand. This is consistent with the historical evidence that economic prohibition is not a constitutive feature of religious cultures but appears eventually in their evolution. Prohibition against secular sciences should then not be understood as inherent to religious philosophies, but (ironically) as a "doctrinal innovation" that allows strong religious preferences to take root in the population.

Relative to the existence of a timing in the instigation of prohibition, Chaney (2016) shows that a significant drop in books written on scientific topics by authors with an Islamic-sounding name becomes statistically significant in the twelfth century.[12] The author gives some evidence that the increase in the political power of religious elites caused the decline in scientific output. Relatedly, regarding prohibition against usury activities in Europe, the Catholic Church was primarily forbidding them to clerics in the 500-1050 period (Reed and Bekar (2003)). Usury became a dominant concern only during the 1175-1350 period, with a peak at the Council of Lyon in 1274 and at the Council of Vienne in 1312 (Reed and Bekar (2003)). The enforcement of the prohibition was relaxed only by 1830, when "the Sacred Penitentiary issued instructions to confessors not to disturb penitents who lend money at the legal rate of interest." (De Roover (1974, p. 321) quoted by Reed and Bekar (2003)). Finally, the theoretical foundations set in this paper suggest that the turning point in the history of a religious culture where economic prohibition appears is not necessarily a period where the discriminated activities particularly threaten the conservative culture (see Section 2.2.3).

12. Chaney (2016) proxies the scientific production in the Muslim pre-industrial world with Harvard's library holdings.

2.3.4 Collusion between political and religious authorities

In this section, I extend the preceding framework in order to account for the existence of collusion equilibria between political and religious elites. Indeed, I have assumed so far that the religious leader was not constrained when collecting contributions from the population or when implementing economic prohibition. However, it can reasonably be argued that political rulers are not neutral when religious authorities seek to collect revenues from the population or intend to manipulate labor outcomes. In the case of Muslim countries for instance, it has been argued that it is precisely when the conservative religious leaders gained political leadership in the prevailing institutional equilibrium that the scientific production declined (Chaney (2016)).

A central incentive of political elites to collude with religious leaders is their willingness to gain legitimacy. This argument has been advanced in the literature by Coşgel and Miceli (2009), Rubin (2011) or Auriol and Platteau (2017) among others.[13] I modify the model of the preceding section by assuming that there is a political authority that is a first mover and that can decide whether it will collude with the religious leader. I make the legitimacy-seeking nature of collusion stark by assuming that when there is no collusion, the political ruler faces a positive probability of being overthrown - which depends positively on the level of extraction - while if he colludes, he faces no threat of social unrest. Furthermore, I posit that if there is collusion, then the political leader delegates the provision of public goods to the religious authority, who is not constrained in implementing economic prohibition.

Let η_0 be the share of the surplus made by the religious authority that is redistributed to the political authority when there is collusion. The timing of the game unfolds as follows:

— The religious authority offers a share η_0 of its surplus to the political authority if he accepts to collude.

— If the offer is not accepted, then the political authority extracts a fraction T_t of the tax base as rents in period $t \geq 0$. The religious authority can neither collect contributions

13. Auriol and Platteau (2017) give few examples of the working of co-optation of religious elites by political rulers and see in the lack of centralization of Islamic religious authorities a key impediment to progressive reforms, as opposed to the more centralized structure of Christianity.

on the citizenry nor impose economic prohibition.

— If the offer is accepted, then the political leader delegates the provision of public goods as well as the collection of individual contributions to the religious leader, who is not constrained in implementing economic prohibition.

When there is no collusion, the political authority extracts revenues from the population, leaving only few resources left for the agents to contribute to the religious good. I assume that the religious authority can not collect contributions in this case for simplicity. Furthermore, the political authority does not let the religious leader impose economic prohibition on the population. This could be explained in a more complex model by the existence of some positive benefits for the political leader from having a diversified economy (or a more straightforward direct interest in scientific discoveries and/or usury). Including this complexity will not change the result.

When an alliance is made between the political and the religious authorities, the latter collects individual contributions and can freely impose economic prohibition on the citizenry. The joint surplus of the two leaders equalizes the surplus of the religious authority, who redistribute a share η_0 to the political authority. When economic prohibition is implemented, the cost c of enforcing it can either be paid by the religious authority or by the political ruler without loss of generality.

No collusion case: I assume that when there is no collusion, the population rebels and overthrows the political leader with some positive probability $f(T_t)$, which depends positively on the level of expropriation $T_t \in [0,1]$. The function $f(.)$ is increasing and convex, with $f(0) = 0$ and $f(1) < 1$. I assume that the time preference of the political (resp. religious) authority is equal to ρ_p (resp. ρ_r). When there is no alliance, the political authority sets $\{T_t^*\}_{t\geq 0}$ such that:

$$\{T_t^*\}_{t\geq 0} = \arg\max_{\{T_t\}} \int_0^\infty e^{-\rho_p t}(1 - f(T_t))T_t w_1(y_0)dt,$$

since the tax base is equal to the size of the population multiplied by the individual revenues

$w_1(y_0) = w_2(y_0)$. I assume that $w_1(y_0) = 1$ in the sequel. It follows immediately that in any period t, $T_t^* = T^*$, with

$$1 - f(T^*) - f'(T^*)T^* = 0$$

when the solution is interior, which I assume for simplicity.[14] Consequently, the utility of the political ruler is equal to

$$W^{NC} = \frac{1}{\rho_p}(1 - f(T^*))T^*$$

when there is no collusion.

Collusion case: The religious authority maximizes its surplus and offers a share η_0 to the political authority at the beginning of the game. The religious authority maximization program can be written in the form

$$\begin{aligned} &\max_{\eta_0, \tilde{t}} W(\tilde{t}, \rho_r) \\ &\text{s.t.} \begin{cases} \eta_0 W(\tilde{t}, \rho_p) \geq W^{NC} \\ \dot{q}_t = \lambda q_t(1 - q_t)(\tau^C(q_t) - \tau^M(q_t)) \\ q(0) = q_0, \end{cases} \end{aligned} \tag{2.22}$$

with $W(\tilde{t}, \rho)$ given by (2.20). Observe that one key difference between (2.22) and (2.19) is that now the maximization is constrained by the feasibility of an alliance with the political authority. Furthermore, since the religious and the political authorities have different time preferences, they will have different valuations for the benefits of an alliance. This will constrain the strategy of the religious authority, as I demonstrate in the next proposition.

Proposition 5. *When the cost of economic prohibition is sufficiently low, for $q_U < q(0) < q^*$ and given that the conditions in Proposition 4 are respected:*

— *Collusion equilibria exist. In a collusion equilibrium, the religious leader redistributes a share $\eta_0 = W^{NC}/W(\tilde{t}^*, \rho_r)$ of its surplus to the political authority, with $\tilde{t}^*$ the unique solution of (2.22).*

14. Given that $f(.)$ is convex, the second-order condition is fulfilled.

— *The likelihood of a collusion equilibrium is higher when the preferences for the religious good increase (i.e. γ_p increases, for $p \in \{M, C\}$), when the provision of the religious good g increases, or when the state capacity decreases (i.e. the function $f(.)$ becomes steeper), when the reservation utility u decreases and when the cost of economic prohibition c decreases.*

— *The duration of economic prohibition increases with the time preferences of both the religious and the political authorities (i.e. it decreases with both ρ_r and ρ_p).*

Proposition 7 is valid under the conditions that guarantee the uniqueness of the interior steady state of the dynamics of religious preferences, given that c is sufficiently low (see Proposition 4). The determination of a collusion equilibrium is represented in Figure 2.5. As represented on the figure, when W^{NC} is sufficiently low - and given that ρ_p and ρ_r are not too dissimilar, $W(\tilde{t}^*, \rho_p)$, the maximum surplus of the religious authority - as evaluated by the political authority - is above W^{NC}. A collusion equilibrium is then feasible. By redistributing a share $\eta_0 = W^{NC}/W(\tilde{t}^*, \rho_r)$ of the surplus to the political authority, the religious leader is certain that his offer will be accepted at the beginning of the game. Thus, the religious will set the policy that represents the minimum deviation from his unconstrained optimum $\tilde{t}_r^*$ and that is such that the collusion constraint binds.

Observe from figure 2.5 that collusion will not be an outcome if the religious leader does not implement discrimination in equilibrium, as $W(0, \rho_p) < W^{NC}$. By contrast, if prohibition

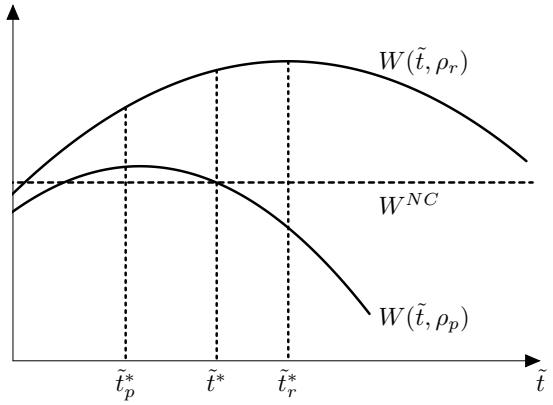

Figure 2.5 – Collusion equilibrium.

is implemented, then the religious leader can offer an better contract to the political authority (i.e. he offers him at least W^{NC}). As discussed in Section 2.2.3 then, this could explain why traditionalist religious leaders gained significant leadership during the Sunni Revival, winning *de facto* the epistemological debate with the rationalists.

From figure 2.5, any factor that increases $W(., \rho_p)$ (resp. decreases W^{NC}) increases the likelihood of a collusion equilibrium. For instance, if the preferences of the conservatives for the religious good γ_C increase, then the curves $W(., \rho_k)$ are shifted up for $k \in \{p, r\}$, because the conservatives invest more in socialization and thus their population grows faster. [15] The reasoning is the same for the rest of the parameters. Thus, any proximate cause of economic prohibition is also a proximate cause of collusion between political and religious authorities.

Besides, it is reasonable to expect political authorities to have lower time preferences than religious authorities. One reason is that political authorities typically stay in power for limited durations, and care about the rents they capture during their incumbency. Thus, we should expect ρ_p to be lower than ρ_r. The consequence of this is that the lower ρ_p relative to ρ_r, then the lower the duration of economic prohibition in a collusion equilibrium. If the difference is sufficiently high, then collusion may not be an equilibrium outcome. This effect is represented in Figure 2.6. Relative to Figure 2.5, $W(., \rho_p)$ is shifted downwardly because the political leader is not willing to pay the cost of economic prohibition for a long period of time, given that he does not value future streams of revenues as much as the religious authority. This result suggests that in economies where the durations of political incumbency are low or volatile, a higher separation of power between political and religious authorities should be observed. Consequently, the existence of "classical" institutional equilibria during the pre-industrial era where religious authorities protect the citizenry from expropriation and control coordination networks while there is a lack of separation between the political and the religious powers is conditioned by a sufficiently high level of political stability (which, in turn, might be guaranteed by religious authorities' ability to control coordination networks).

When the fiscal capacity is weak in that the political authority is unable to extract revenues

15. This is true given that $q(0)$ is initially such that prohibition can be optimal, see Proposition 5.

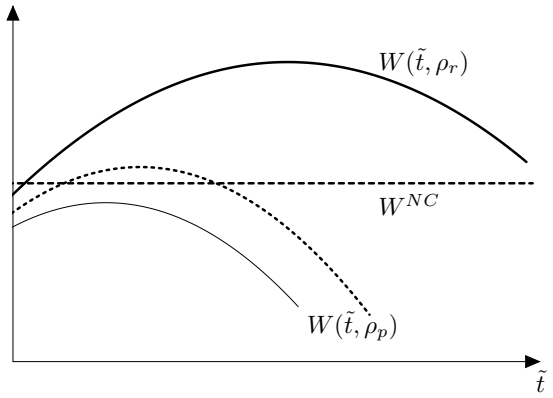

Figure 2.6 – Case with no collusion equilibrium.

from the population without increasing significantly the probability of being overthrown (i.e. $f(.)$ has a steeper profile), then collusion becomes more likely in equilibrium because the gains from being endorsed by religious authorities are high. The consequence of the instigation of such an institutional arrangement between the political and the religious authorities is, of course, that the latter is not constrained in molding the preferences of the population, and in particular, he can implement economic prohibition. Not surprisingly, it means that the stronger the ability of religious leaders to coordinate popular masses, the higher the likelihood of collusion between politics and religion, and the higher the duration of economic prohibition.

One major institutional difference during the middle ages between European countries and Muslim countries is that the former relied on feudal arrangements for raising armies, while the latter used slave armies (Mamlukism). According to Chaney (2011), Mamlukism allowed religious leaders in Muslim countries to maintain a monopoly over popular coordination networks, while in Europe, feudalism led to an increasing separation of power. Consistently with this view, Blaydes and Chaney (2013) give some evidence that the relatively higher executive constraints inherent to the feudal system allowed for significantly longer incumbencies in Europe. Although those considerations are beyond the scope of the model, observe that any institutional arrangement that reduces the incentive of a political ruler to collude with a religious authority will decrease the duration of economic prohibition. Thus, if the rise of the Feudal system in Europe meant that religious leaders lost to some extent their control

over popular coordination networks, then our theory is consistent with a significantly lower duration of economic prohibition in Europe.

Observe that we have abstracted from intrinsic interests of political rulers from having a society that is culturally *and* economically segmented. However, such intrinsic interests may exist when the cultural groups composing the population have different willingness to rebel and political leaders can discriminate rent extraction on a sectoral basis. Indeed, if a minority culture has no mean of rebelling efficiently, then the political ruler could delegate the provision of public goods to the authorities of the main religious denomination and keeps on expropriating minority individuals given that they specialize in one particular type of economic activities. [16]

2.4 Conclusion

In this paper, I have presented a theory that seeks to explain why religions have repeatedly prohibited some economic activities throughout history. I suggested that the existence of economic prohibition is linked to strategic concerns of religious leaders' internalizing the dynamic interaction between the division of labor and intergenerational cultural transmission strategies.

The theory yields two main predictions. First, during transitory periods in their evolution, economic prohibition allows religious norms to diffuse in the population because it creates a *cultural* division of labor. Economic prohibition is then strategically implemented by religious leaders eager to increase the size of their cultural group under specific conditions discussed in the main text. Second, the conditions that favor the implementation of economic prohibition are also key determinants of the existence of collusion equilibria between politics and religion. Thus, the intricate linkage between cultural evolution and occupational decisions may explain both why economic discrimination allows strong religious preferences to take root in the

16. This mechanism is reminiscent of strategies in European countries during the middle ages where jewish minorities were active in financial activities and targeted by high level of extractions where they were authorized to settle.

population and why conservative religious leaders can gain significant political leadership. We discussed this result in the context of the Sunni Revival in Section 2.2.3.

The framework of this paper could be extended so as to account for labor mobility of innovators in a fragmented polity. Indeed, as argued by Mokyr (2016), the migrations of innovators across European countries partly explain the failure of the Church to decrease the rate of scientific discoveries in the region. [17] I have left aside the role of institutions affecting occupational decisions in this paper, thereby abstracting from the rich interactions between guilds or universities, technical change and religion. [18]

Another interesting venue for future research would be to account for state dependency in technology adoption. This, ultimately, could allow for a careful analysis of the long-term consequences on growth and technology adoption of past episodes of economic prohibition. [19] Finally, I have been concerned in this paper with one religious denomination. Yet a large literature has put forth the role of competition on the religious market as an important determinant of religiosity. [20] Such an extension could also help deepen our understanding of the interaction between the diffusion of the Reform and the evolution of occupational decisions in Europe.

17. See as well David (2008) on the linkage between competition among Europe's noble patrons motivated to attract prestigious intellectuals and the openness of scientific knowledge.

18. Richardson and McBride (2009) suggest that guilds relied on religion to sustain cooperation when mortality rates were high. They suggest that in the sixteenth century, when the disease environment eased - and given the availability of a new religious doctrine with the Reformation - new methods or organizing industry could develop. Relatedly, de la Croix, Doepke, and Mokyr (2016) put forth the role of guilds in the development trajectory of Europe. Huff (2003) argues that the rise of universities is an important institutional turning point in the history of European science.

19. Bezin (2016) considers state dependency in technology adoption in a model of cultural evolution.

20. See for instance Finke and Iannaccone (1993), Stark and Iannaccone (1994), Stark and Bainbridge (1996), Barros and Garoupa (2002), Montgomery (2003), McBride (2008) and Montgomery (2010). See as well the review of Iyer (2016). Relatedly, Verdier and Zenou (2015) provide an extension of their dynamic theory of cultural evolution that accounts for competition between several cultural leaders.

2.5 Appendix

2.5.1 Competitive equilibrium

The market clearing conditions on the markets for good 1 and good 2 imply that the relative price of good 2 p has to satisfy

$$p \equiv \frac{p_2}{p_1} = \frac{1-\alpha}{\alpha}\left(\frac{Y_2}{Y_1}\right)^{-\frac{1}{\sigma}}. \tag{2.23}$$

The response of the relative price to the relative supply depends on the elasticity of substitution σ, and the greater the relative supply of good 2, the lower is its relative price p.

We denote π_i the shadow price of sector i-specific technologies, given that the production of sector-specific technologies is assumed to be monopolistic.[21] The first-order conditions for the maximization of profits in sector i are

$$\begin{cases} p_i \beta \left(\frac{y_i}{a_i}\right)^{1-\beta} = \pi_i \\ p_i (1-\beta)\left(\frac{a_i}{y_i}\right)^{\beta} = w_i \end{cases} \tag{2.24}$$

with $y_2 = y$ and $y_1 = 1 - y$. Thus, the LHS of the first equation above gives a downward slopping inverse demand function for sector i-specific technologies $\pi_i(a_i)$. The profits of a monopoly supplying sector i-specific technologies can be written as $Z_i(a_i) = (\pi_i(a_i) - c)a_i$ where c denotes a constant marginal cost (that we take equal in both sectors for simplicity). Following Acemoglu (2002), we deduce that the price of sector i-specific technologies is set to

$$\pi_i(a_i) = \frac{c}{\beta}, \tag{2.25}$$

meaning that the shadow price of technologies is independent from sector-specific parameters (which comes from the iso-elasticity of the demand curves, as observed by the preceding

21. This assumption says no more than patents or adaptation delays make innovations exclusive to one producer for at least one period of time.

author). We assume that $c = \beta$ in the rest of the paper for simplicity, so the shadow-price of technologies is necessarily equal to 1 in both sectors. Thus, from (3.7), we deduce that the demand for sector i-specific technologies is such that

$$a_i = y_i \beta^{\frac{1}{1-\beta}} p_i^{\frac{1}{1-\beta}}, \tag{2.26}$$

so from the first-order conditions, the wage in sector i is given by

$$w_i = (1-\beta)\beta^{\frac{\beta}{1-\beta}} p_i^{\frac{1}{1-\beta}}. \tag{2.27}$$

Substituting the demand for technologies (9) in the production functions (6), we get

$$\frac{Y_2}{Y_1} = \frac{y}{1-y}\left(\frac{p_2}{p_1}\right)^{\frac{\beta}{1-\beta}}. \tag{2.28}$$

Finally, substituting the relative production in (3.9) with (3.5), we obtain the relative price of the two goods as a function of y, the labor demand in sector 2,

$$p = \left[\left(\frac{1-\alpha}{\alpha}\right)^{\sigma}\left(\frac{1-y}{y}\right)\right]^{1/\epsilon} \tag{2.29}$$

where $\epsilon = (\beta + \sigma(1-\beta))/(1-\beta)$. From that point, we deduce that the relative wage in sector 2 must verify

$$\frac{w_2}{w_1} = \left[\left(\frac{1-\alpha}{\alpha}\right)^{\sigma}\left(\frac{1-y}{y}\right)\right]^{1/\epsilon(1-\beta)} \tag{2.30}$$

and the relative production of sector 2-specific technologies is such that

$$\frac{a_2}{a_1} = \frac{y}{1-y} p^{\frac{1}{1-\beta}}. \tag{2.31}$$

The wage premium in sector 2 decreases with the size of that sector i.e. with y. Moreover, there exists a tension between a price and a market effect, consistently with Acemoglu (2002). Indeed, the technology production is biased toward the sector producing the most expensive good

(price effect) *ceteris paribus*. At the same time, since technology and labor are complement in the production of the two outputs, a decrease in the labor supply in sector i y_i will shift downwardly the demand for technologies in sector i, thereby affecting negatively the relative technical coefficient that sector (market size effect). From (2.29), we deduce that

$$\frac{a_2}{a_1} = [(\frac{y}{1-y})^{(1-\beta)(\sigma-1)}(\frac{1-\alpha}{\alpha})^{\sigma}]^{1/\epsilon(1-\beta)}, \tag{2.32}$$

meaning that the relative technical coefficient in sector 2 increases with y, the size of sector 2 given that the elasticity of substitution between the two goods σ is above unity (i.e. good 1 and good 2 are gross substitutes).

2.5.2 The dynamics of religious preferences

Following Bisin and Verdier (2000b, 2001), I assume that preferences are acquired during childhood. The parents control the exposition to role models of their naive offspring, which are either matched with their family ("vertical transmission"), or in the population at large ("oblique transmission"). [22] Parents are assumed imperfectly empathic in their transmission strategy and they perceive the welfare of their children through their own utility function. A Poisson birth and death process of parameter $\lambda < 1$ is assumed in each period t, so that the population size is kept constant.

I denote the effort of socialization τ_t^i in period t for $i \in \{C, M\}$. Given Proposition 2, there are four different cases to consider, depending on the allocation on the labor market. I denote q_{t+1}^e the expected value of the fraction of conservatives in period $t+1$. The transition

22. This terminology is consistent with the literature on cultural transmission, see Boyd and Richerson (1985), Cavalli-Sforza and Feldman (1985) for example.

probabilities P_t^{ij} that a i-type parent has a j-type child are given by

$$\begin{cases} P_t^{CC} = \tau_t^C + (1-\tau_t^C)(1-q_t) \\ P_t^{CM} = (1-\tau_t^C)(1-q_t) \\ P_t^{MM} = \tau_t^M + (1-\tau_t^M)q_t \\ P_t^{MC} = (1-\tau_t^M)q_t. \end{cases} \tag{2.33}$$

The effort τ_t^i that parents of type i will devote to cultural transmission solves

$$\tau_t^i = \arg\max_{\tau^i \geq 0} -H(\tau) + P_t^{ii} u^{ii}(q_{t+1}^e) + P_t^{ij} u^{ij}(q_{t+1}^e) \tag{2.34}$$

with $H(.)$ an increasing convex function such that $H(0) = 0$ and $H'(0) = 0$. The function $u^{ij}(.)$ gives the utility of being a type $j \neq i$, as evaluated by a parent of type i under the imperfect empathy assumption and the function $u^{ii} \equiv u^i(.)$ is given in (3.21). I assume that $q_{t+1}^e = q_t$. In this case, the utility $u^{ij}(q_t)$ will depend on occupational choices of the different cultural groups as well as on the sorting effect. Assuming that $H(\tau) = \tau^2/2$, we get the standard result

$$\begin{cases} \tau^C(q_t) = (1-q_t)(u^{CC}(q_t) - u^{CM}(q_t)) \\ \tau^M(q_t) = q_t(u^{MM}(q_t) - u^{MC}(q_t)) \end{cases} \tag{2.35}$$

from the maximization (3.23). Finally, during a time lapse dt a fraction $\lambda dt(1-q_t)q_t(1-\tau^M(q_t))$ (resp. $\lambda dt(1-q_t)q_t(1-\tau^C(q_t))$) of moderate (resp. conservative) children are socialized by conservatives (resp. moderates). Thus, the cultural dynamics is such that

$$q_{t+dt} - q_t = \lambda dt[(1-q_t)q_t(1-\tau^M(q_t)) - (1-q_t)q_t(1-\tau^C(q_t))], \tag{2.36}$$

which rewrites

$$\dot{q}_t = \lambda q_t(1-q_t)(\tau^C(q_t) - \tau^M(q_t)). \tag{2.37}$$

2.5.3 Proof of Proposition 2

We denote $v^{ij,k}(q_t)$ the utility evaluated by an agent of type $i \in \{C, M\}$ from being an agent of type $j \in \{C, M\}$ working in sector $k \in \{1, 2\}$.

Case 1: $q_t < q_C$. Conservatives have no influence on occupational choices so there is no conflict over the optimal allocation of labor. Both the conservatives and the moderates are indifferent between the two types of activities and $y = y_0$. Moreover, conservatives are not sufficiently numerous to enforce contributions. Parents have no incentive to transmit their own cultural trait, $\tau_t^C = \tau_t^M = 0$ and $\dot{q}_t = 0$.

Case 2.1: $y(q_t) = y_M(q_t)$ **and** $q_t > q_M$. In this case, we know from Proposition 2 that there is a partial sorting on the labor market where conservatives only populate sector 1 while there is a positive fraction of moderates in both sectors. Moreover, while working in sector 1, agents contribute independently from their cultural type since $q_t > q_M > q_C$. Thus, if a conservative parent is not successful in socializing his offspring, he knows that his kid will have a positive probability of choosing an occupation in sector 2. But the utility of having an occupation in sector 2 - from the perspective of a conservative - is strictly lower from that of working in sector 1 given that $y(q_t) = y_M(q_t) > y_C(q_t)$. Thus, when $y(q_t) = y_M(q_t)$ and $q_t > q_M$,

$$\begin{cases} u^{CC}(q_t) = v^{CC,1}(q_t) \\ u^{CM}(q_t) = \frac{y_M(q_t)}{1-q_t} v^{CM,2}(q_t) + \frac{1-q_t-y_M(q_t)}{1-q_t} v^{CM,1}(q_t) \end{cases} \tag{2.38}$$

since a fraction $y_M(q_t)/1 - q_t$ (resp. $1 - q_t - y_M(q_t)/1 - q_t$) of moderates works in sector 2 (resp. 1). From (3.11),

$$\begin{cases} v^{CC,1}(q_t) = v^{CM,1}(q_t) = w_1(y_M(q_t))(1-e)[\gamma_C g + u] \\ v^{CM,2}(q_t) = w_2(y_M(q_t))[(1-q_t)\gamma_C g + u] \end{cases} \tag{2.39}$$

where the wages are at their equilibrium values so they depend only on the labor allocation $y(q_t)$. Similarly for the moderates, since $y(q_t) = y_M(q_t)$, they are indifferent between the two

sectors and all the conservatives are in sector 1 (Proposition 2),

$$\begin{cases} u^{MM}(q_t) = v^{MM,2}(q_t) = v^{MM,1}(q_t) \\ u^{MC}(q_t) = v^{MC,1}(q_t) = v^{MM,1}(q_t) \end{cases} \tag{2.40}$$

with

$$\begin{cases} v^{MM,1}(q_t) = v^{MC,1}(q_t) = w_1(y_M(q_t))(1-e)[\gamma_M g + u] \\ v^{MM,2}(q_t) = w_2(y_M(q_t))[(1-q_t)\gamma_M g + u]. \end{cases} \tag{2.41}$$

Substituting (3.32) and (2.41) in (3.24), we deduce the efforts of transmission as functions of q_t,

$$\begin{cases} \tau^C(q_t) = y_M(q_t)(w_1(y_M(q_t))(1-e)[\gamma_C g + u] - w_2(y_M(q_t))[(1-q_t)\gamma_C g + u]) \\ \tau^M(q_t) = 0 \end{cases} \tag{2.42}$$

so it is immediate that $\dot{q}_t > 0$ in such a situation. Religious strive when they induce a distortion on the labor allocation for $q_t > q_M$ that makes *moderates* indifferent between the two sectors i.e. such that $y(q_t) = y_M(q_t)$ (the religious do so by shifting the wage premium in sector 2 above unity and thus by affecting negatively the demand for labor in sector 2, see the preceding section). Since moderates are indifferent between the two sectors, there is a sorting effect where conservatives only populate sector 1 which grants strictly higher utility levels from their perspective. This makes them invest in inter-generational transmission, while moderates are indifferent so they do not invest in transmitting their cultural trait.

Case 2.2: $y(q_t) = y_M(q_t) = y_0$ **and** $q_C < q_t < q_M$. Since $q_C < q_t < q_M$, conservatives and moderates adopt different contributing behaviors while working in sector 1. That is, conservatives expect that if socialization fails, their offspring will stop contributing to the provision of the religious good, which is suboptimal from their perspective since $q_C < q_t$. Additionally, observe that moderate parents expect that after a failed socialization, their

offspring will start contributing which is necessarily suboptimal since $q_t < q_M$. Thus,

$$\begin{cases} v^{CM,1}(q_t) = w_1(y_M(q_t))(1-e)[(1-q_t)\gamma_C g + u] \\ v^{CC,1}(q_t) = w_1(y_M(q_t))(1-e)[\gamma_C g + u] \\ v^{CC,2}(q_t) = v^{CM,2}(q_t) = w_2(y_M(q_t))[(1-q_t)\gamma_C g + u] \end{cases} \tag{2.43}$$

As $w_1(y_M(q_t)) = w_2(y_M(q_t))$ for $y(q_t) = y_M(q_t) = y_0$ (see Proposition 2), there is no wage premium at working in sector 2, $v^{CM,1}(q_t) = v^{CM,2}(q_t)$.

$$\begin{cases} u^{MM,2}(q_t) = v^{MC,2}(q_t) = w_2(y_M(q_t))[(1-q_t)\gamma_M g + u] \\ u^{MM,1}(q_t) = w_1(y_M(q_t))[(1-q_t)\gamma_M g + u] \\ u^{MC,1}(q_t) = w_1(y_M(q_t))(1-e)[\gamma_M g + u]. \end{cases} \tag{2.44}$$

We deduce the efforts of transmission,

$$\begin{cases} \tau^C(q_t) = (1-q_t)w_1(y_M(q_t))((1-e)(\gamma_C g + u) - ((1-q_t)\gamma_C g + u)) \\ \tau^M(q_t) = q_t w_1(y_M(q_t))(((1-q_t)\gamma_M g + u) - (1-e)(\gamma_M g + u)). \end{cases} \tag{2.45}$$

We deduce that

$$\Delta\tau^C(q)/w_1(y_M(q)) = (1-q)\{(1-e)(\gamma_C g+u)-((1-q)\gamma_C g+u)\}-q\{(1-q)\gamma_M g+u-(1-e)(\gamma_M g+u)\}, \tag{2.46}$$

which rewrites

$$\Delta\tau^C(q)/w_1(y_M(q)) = (q-e)(\gamma_C(1-q)+\gamma_M q)g - eu. \tag{2.47}$$

Since $\Delta\tau^C(.)$ is a strictly concave function of q and $\Delta\tau^C(q_C) < 0$ and $\Delta\tau^C(q_M) > 0$, we deduce that $\Delta\tau^C(q) = 0$ admits a unique solution. Thus, there exists only one unstable rest point of the cultural dynamics in the region $q \in (q_C, q_M)$.

Case 3: $y(q_t) = y_C(q_t)$. In this case, the conservatives populate both sectors, while the moderates are present only in sector 2. By the same token as before, since the conservatives are at an indifference point, we find $\tau_t^C = 0$. Observe that since all the moderates work in sector 2 when $y(q_t) = y_C(q_t)$, we need not to distinguish between the cases $q_t < q_M$ and $q_t > q_M$. By contrast with the conservatives, the moderates will invest a positive effort in inter-generational transmission. Indeed, if vertical transmission fails they expect their offspring to choose an occupation in sector 1 with a strictly positive probability. Formally,

$$\begin{cases} v^{MM,2}(q_t) = v^{MC,2}(q_t) = w_2(y_C(q_t))[(1-q_t)\gamma_M g + u] \\ v^{MM,1}(q_t) = w_1(y_C(q_t))[(1-q_t)\gamma_M g + u] \\ v^{MC,1}(q_t) = w_1(y_C(q_t))[(1-q_t)\gamma_M g + u] \text{ if } q_t < q_M \\ v^{MC,1}(q_t) = w_1(y_M(q_t))(1-e)[\gamma_M g + u] \text{ if } q_t > q_M, \end{cases} \tag{2.48}$$

which implies that $\tau_t^M > 0$ whenever $y(q_t) = y_C(q_t)$, so $\dot{q}_t < 0$ necessarily.

Case 4: $y(q_t) = 1 - q$. This case corresponds to a perfect sorting on the labor market. By the arguments developed above, both cultural types will devote a positive effort to cultural transmission that may account for both the cultural and economic motives when $q_t < q_M$ and only for the economic motive when $q_t > q_M$. Following the same reasoning as before, we find that

$$\begin{cases} \tau^C(q_t) = (1-q_t)(w_1(1-e)(\gamma_C g + u) - w_2[(1-q_t)\gamma_C g + u]) \\ \tau^M(q_t) = (q_t(w_2[(1-q_t)\gamma_M g + u] - w_1(1-e)(\gamma_M g + u)). \end{cases} \tag{2.49}$$

Consequently, $\Delta\tau^C(q_t) = (\tau^C(q_t) - \tau^M(q_t))/w_1$ writes

$$\Delta\tau^C(q_t) = (1-e)[g\{(1-q_t)\gamma_C + q_t\gamma_M\} + u] - \eta(q_t)\{u + g[(1-q_t)^2\gamma_C + q_t(1-q_t)\gamma_M]\}. \tag{2.50}$$

2.5.4 Proof of Proposition 3

The second term in the RHS of (44) is decreasing if

$$f(q) = \eta(q)\{u + g[(1-q)^2\gamma_C + q(1-q)\gamma_M]\} \tag{2.51}$$

increases with q. Given the expression of the wage premium in (13) for $y(q) = 1 - q$, we deduce that $f'(q) > 0$ when

$$\eta'(q)/\eta(q) > \frac{g[2(1-q)\gamma_C - (1-2q)\gamma_M]}{u + g((1-q)^2\gamma_C + q(1-q)\gamma_M}. \tag{2.52}$$

Since

$$\eta'(q)/\eta(q) = \frac{1}{Z}\frac{1}{q(1-q)}, \tag{2.53}$$

with $Z = \beta + \sigma(1-\beta)$, we deduce that $f'(q) > 0$ if

$$\frac{1}{Z}\frac{1}{q(1-q)} > \frac{g[2(1-q)\gamma_C - (1-2q)\gamma_M]}{u + g((1-q)^2\gamma_C + q(1-q)\gamma_M}. \tag{2.54}$$

As (i) $q(1-q) \leq 1/4$ (ii) the denominator in the RHS is minimum for $q = 0$ and equalizes $\gamma_C g + u$ and (iii) the numerator in the RHS decreases with q, is maximum for $q = 0$ and equalizes $g(2\gamma_C - \gamma_M)$, a sufficient condition for the last equation to hold is that

$$\frac{1}{Z} > 4\frac{g(2\gamma_C - \gamma_M)}{u + \gamma_C g}, \tag{2.55}$$

which establishes the sufficient condition on σ, given that $Z = \beta + \sigma(1-\beta)$. Observe that the RHS is decreasing with u, so the condition is necessarily fulfilled when the agents have a sufficiently large reservation utility.

2.5.5 Proof of Proposition 4

Claim 2.1. *It is never optimal for the religious leader to pay the cost c of discriminating economic occupations when $q_C < q_t < q_M$.*

Démonstration. Notice that in the region where $q_C < q_t < q_M$, the dynamics is the same whether or not economic discriminations exist. This is because moderate will not contribute anyway (since $q_t < q_M$), so they get the same punishment whether they work in sector 1 or in sector 2. This implies that moderates will be indifferent between the two sectors for $y_M(q_t) = y_0$ (and for a wage premium equal to 1). But conservatives adopt different behaviors in the two sectors. Since $q_C < q_t$, they contribute while working in sector 1 and stop doing so while working in sector 2. Consequently, they will ask for a wage premium above unity in sector 2, meaning that $y_C(q_t) < y_0$. As labor decisions are not made cooperatively, $y(q_t) = y_0$ necessarily, so the labor market is not affected by economic discriminations. This explains why it is never optimal to pay the cost c of discriminating economic occupations when $q_C < q_t < q_M$. □

It is direct that the first-order condition associated with the maximization program (2.19) is given by (2.21). We have represented $\frac{W}{\rho}\dot{q}(\tilde{t} + t_b(q_0)) - c$ as a function of $q(\tilde{t} + t_b(q_0))$ in figure 2.7. If c is sufficiently low, then the red curve is sufficiently close from the curve $q(\tilde{t} + t_b(q_0)) \to W/\rho\dot{q}(\tilde{t} + t_b(q_0))$, which crosses only once the horizontal axe under the uniqueness condition established in Proposition 4. If c is high, given that the monotonicity of $q(\tilde{t} + t_b(q_0)) \to \dot{q}(\tilde{t} + t_b(q_0))$ is not guaranteed, then even though $q(\tilde{t} + t_b(q_0)) \to \dot{q}(\tilde{t} + t_b(q_0))$ crosses only one the horizontal axe under the condition of Proposition 4, it may be that $q(\tilde{t} + t_b(q_0)) \to \frac{W}{\rho}\dot{q}(\tilde{t} + t_b(q_0)) - c$ crosses it multiple times.

Observe finally from figure 2.6 that the first-order condition (2.21) holds with equality whenever $q(\tilde{t} + t_b(q_0)) < q^*$.

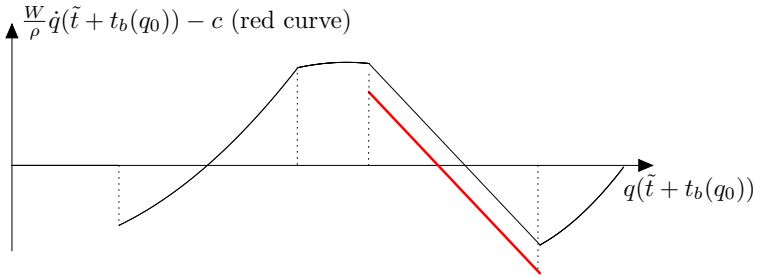

Figure 2.7 – In red: $\frac{W}{\rho}\dot{q}-c$ as a function of q.

Chapter 3

Multi-candidate Political Competition and the Industrial Organization of Politics[1]

1. This chapter is based on a joint work with Thierry Verdier. We are grateful to Gani Aldashev, Alberto Bisin, Laurent Bouton, Micael Castanheira, Rachel Kranton and Ragnar Torvik for their discussions and insightful comments that greatly improved the paper. We thank the participants to the ECARES and PSI-PSE seminars. Avner Seror and Thierry Verdier acknowledge the financial support from the ERC Grant TECTACOM n° 324004.

Abstract

In this paper, we present and micro-found a theory of multi-candidate political competitions that allows to study various issues related to the industrial organization of politics. The flexibility of our analytical framework is demonstrated through several applications on the topics of special interest politics, coalition formation in the legislature in proportional elections, redistribution under alternative electoral rules and franchise extension.

JEL codes: D71, D72, L11.

Keywords: Political Economy, Elections, Probabilistic Voting Models, Fréchet distributions, Industrial Organization, Redistribution, Public policy, Duverger, Run-off, Plurality, Proportional, Franchise, Democracy.

3.1 Introduction

Political competition is more than often modelled by political economists as a two-party contest for power. This is for instance examplified by the celebrated Hotelling model of deterministic voting, as well as the well known probabilistic voting model applied to so successfully to different economic contexts (Persson and Tabellini (2002)). Yet, many issues in political economy involve the study of multiple candidates' competition. As such, one finds for instance the political fragmentation under alternative electoral rules, the formation of coalitions in the legislature in proportional systems, and more broadly any topic that considers the endogenous entry of participants in political competitions. While some analytical frameworks of multi-candidates competition have been developed to tackle these questions, in most cases these models are not easy to handle. On the one hand, they may face problems of existence of equilibria (deterministic spatial voting models), or on the contrary they may involve many equilibria (citizen-candidates models), or the applications are exogenously limited to a few number of candidates, typically three or four parties, highly dependent on the institutional and political context under scrutiny.

The purpose of this paper is to provide a simple analytical framework on multi-candidate elections that is tractable enough to allow a systematic analysis of the endogenous structure of political competition in ways similar to the Industrial Organization models used to analyse endogenous market structures under economic competition. The versatility of our approach is demonstrated through several applications on classical topics in political economy: special interest politics on redistribution and public good provision, coalition formation in the legislature, and franchise extension in the democracy.

The model that we propose in the second section of this paper is a standard probabilistic voting theory (e.g. Coughlin (1992) and Persson and Tabellini (2002)). We consider however that the noise in random voting decisions is distributed according to some Fréchet (or extreme type II) distributions. We do so because the maximum of a finite sequence of random variables distributed according to Fréchet distributions is a contest function. This idea has already been

exploited in the context of trade between multiple countries by Eaton and Kortum (2002). It is particularly relevant in the context of political competitions as well, since individuals cast their vote for their most preferred candidate out of a finite list of challengers. Our approach implies a simple formalism for the determination of electoral equilibria and for dealing with the issue of the endogenous entry of candidates in election.

We also outline an axiomatic approach to probabilistic voting models that provides a microfounded rationale for the use of Fréchet distributions in voting theories. The analysis is inspired of the seminal approach of McFadden (1974) of individual choice decisions. We assume that in a probabilistic voting model, voting behaviors follow three Axioms. Those axioms are respectively the Independence of Irrelevant Alternatives, the positivity and the Irrelevant of Alternative Set Axioms (McFadden (1974)). As an illustration of our result, assume that the voters use a random voting model. We seek to understand what kind of distribution for the noise in voting decisions is consistent with the three preceding axioms and find that only the Fréchet distributions satisfy them.

We present several applications of the canonical model. The first application studies the endogenous entry of political parties in the context of special interest politics with homogeneous entrants and heterogeneous voters under plurality rules. We provide a closed form result for the equilibrium number of entrants in a symmetric equilibrium. The second application extends the framework of the fist application to the case of heterogeneous candidates. As a first illustration of the heterogeneous case, we consider the sequential entry of candidates with different popularities or valences. We show that when a high valence party is the initial entrant, then it deters the entry of low valence parties relative to a case where parties are homogeneous. Furthermore, we show that the electoral equilibrium is uniquely determined but is not symmetric, since the high valence party will capture more rents than its challengers. We then study the case where candidates' heterogeneity in election is endogenous. We indeed consider that parties can choose their cost of formation, and that higher investments at the formation stage lead to higher valences or popularity levels. We find that in the unique existing equilibrium, the polity is necessarily heterogeneous, meaning that there will be both

"low-cost" and "high-cost" candidates running for the election. Furthermore, we show that allowing for high investments at the party formation stage leads to a less fragmented polity and to more rent extraction from political parties. Besides suggesting a potential drawback of campaign investments, this application provides a rationale for the existence of small parties in a model a stochastic and sincere voting.

The third application extension studies parliamentary systems with a proportional representation electoral system. Following Baron and Diermeier (2001), we assume that a party in the legislature is selected to form a governing coalition with a probability that equalizes its seat share. Our model allows to study the interaction between public policies, coalition formation in the legislature and party formation before the election. To our knowledge, there is no existing study that accounts for both the endogenous entry of parties in proportional elections and the formation of coalitions in the legislature. We are able to give a closed form result for the number of parties in the legislature. Furthermore, we predict that proportional systems should favor a lower party fragmentation than plurality elections. This is because the likelihood to be the formateur of a governing coalition, or the perks from participating to a governing coalition more broadly are higher when parties are relatively well represented in the legislature. This always outpaces the expected cost of forming coalitions, since those are born by the citizenry in equilibrium.

The fourth application considers the effect of alternative electoral rules on political fragmentation. As formulated by Riker (1982, p. 754), the Duverger's hypothesis states that "the simple majority system with second ballot and proportional representation favors multipartyism". In contrast to the Duverger's hypothesis, we show that a plurality system should lead to *more* entrants than a runoff system. Indeed, in our theory with sincere and stochastic voting, the difference between runoff and plurality elections pins down to the effect of the two electoral systems on the marginal benefits of parties from capturing rents. We show that considered separately, the two rounds of runoff elections create a stronger incentive for parties to capture rents, since they imply a lower degree of competition. But taken together, the two rounds create a weaker incentive for parties to capture rents and this explains why

we should expect less parties in equilibrium to form under the runoff system.

The final application of this theory considers the effect of franchise extension in a democracy when the fragmentation of the polity is endogenous. We show that a franchise extension has two main effects on equilibrium political fragmentation. First we show that the type of individuals touched by an extension of the franchise matters for the equilibrium outcome. More precisely, when the newly enfranchised individuals are more (respectively less) politically reactive than the average of the already enfranchised ones, an extension of the franchise reduces (respectively increases) the equilibrium degree of political fragmentation. Second, in a context in which the state capacity to transfer/tax individuals is imperfect, we uncover a positive income effect on equilibrium political fragmentation that is inherent to the process of franchise extension, independently from the preferences of the citizenry. Moreover because of this effect, it can be consensual for an enfranchised elite to extend the franchise, given that it could disproportionally benefit from it when the newly enfranchised interests are not politically reactive. Furthermore, we suggest that franchise extension should precede fiscal consolidation, since the income effect implied by the extension of the franchise is significantly stronger when the fiscal capacity is weak. We are finally able to establish that franchise extension should be expected in relatively more egalitarian economies, thereby providing some theoretical support to the analysis of Engerman and Sokoloff (2000) and Engerman and Sokoloff (2005).

3.1.1 Literature review

Probabilistic voting theories: Electoral competitions have been a major field of research since the seminal works of Downs (1957) and Black (1958). In this endeavor, probabilistic voting models emerged in the 1970s with the works of Brams and O'Leary (1970), Hinich, Ledyard, and Ordeshook (1972), Hinich (1977) and Coughlin and Nitzan (1981).[2] Probabilistic voting theories are particularly useful in dealing with the multidimensionality of political

2. See as well the excellent review of Coughlin (1992) on the early literature on probabilistic voting. More recent works on probabilistic voting include for example Dixit and Londregan (1996), Lindbeck and Weibull (1987) and Persson and Tabellini (2002).

decisions and electoral equilibria typically exist (see for instance Coughlin (1992) and Banks and Duggan (2005)). Probabilistic voting theories become unfortunately excessively complex in the case of multi-candidate elections, which is why most of the existing studies on the subject focus on two-candidate elections.

There has been few attempts in the literature to study multi-candidate elections with probabilistic voting theories. Lin, Enelow, and Dorussen (1999) provide an existence theorem for electoral equilibria for multi-candidate elections. They assume that voters' utility depend on the distance between their own ideal policy and the winning candidate's platform as well as on a random shock. They show that if the utility shocks have a high enough variance, then the expected vote functions of the candidates are concave, which implies the existence of an electoral equilibrium. Our axiomatic approach allows to link the existence of a global equilibrium to the randomness of voting decisions as well. Consistently with Lin, Enelow, and Dorussen (1999), we find that if voters are not too reactive to political platforms, i.e. if the randomness of voting behaviors is sufficiently large, then an electoral equilibrium exists and is unique.

The closest model of multi-candidate election to our knowledge is Schofield (2007). The author uses Gumbel (extreme type I) distributions to build a model of multi-candidate elections. He notices that this type of distribution respects the "independence of irrelevant alternative property". Our axiomatic approach to probabilistic voting further this intuition and micro-founds the use of extreme type II distributions in stochastic voting models. Schofield (2007) shows that candidates in election do converge to the mean platform when there is no large asymmetry in the electoral perception of the "quality" (or valence) or the parties, given that the variance in voters' ideal positions is not too large. The use of extreme type II distribution significantly ease the issue of the existence of electoral equilibria in the case of plurality elections as we will demonstrate in the main text. Furthermore, we are able to derive closed form results for electoral platforms, so we can directly link the valence of candidates in elections to the convergence of the electoral equilibrium. We show in the second application that heterogeneity in the electoral perception of the "quality" of the candidates necessarily

leads to diverging platforms in the case of plurality elections. Furthermore, the formalism implied by the extreme type II distributions allows to study how candidates' heterogeneity affects the endogenous entry of political parties. [3]

Spatial competition theories and endogenous entry: In spatial voting models, policy options are represented by points on an Euclidian space and each voter's utility function is commonly assumed to be a decreasing function of the Euclidian distance between candidates' positions and voter's ideal point. This literature provides an analytical framework to study the existence and the convergence of electoral equilibria with endogenous entrants. Cohen and Shepsle (1990) provides a review on the early literature on the subject (see as well Shepsle (1991) and Osborne (1995) for overviews). In this literature, the policy motivations of candidates in elections are central in understanding the process of entry. In the citizen-candidate model of Osborne and Slivinski (1996) for instance, a three-candidate equilibrium entails entry by a candidate with no chance of winning. Yet her entry causes the winner to be her favorite out of the other two candidates (see as well Palfrey (1984) on similar mechanisms). Osborne (1993) and Weber (1997) introduces sequentiality in entry decisions in spatial models while Feddersen, Sened, and Wright (1990) accounts for strategic voting. Dickson and Scheve (2010) consider finally a theory of electoral institutions with an endogenous number of candidates in a citizen-candidate framework adapted from Osborne and Slivinski (1996). They incorporate an identity-related political behavior in a model of electoral competition using the framework on identity introduced by Akerlof and Kranton (2000, 2010). In their theory, social identities provide a motivation for political behavior, including vote choice and decisions to seek office. Although we consider no policy motivations for office holders, as Dickson and Scheve (2010), we find that the characteristics of the citizenry can have a profound influence on the fragmentation of the polity in either plurality or runoff elections.

Our theory contrasts with the spatial competition theories in two important ways. First,

3. Few works have estimated probabilistic voting models with extreme type I distributions, see for instance Schofield, Sened, and Nixon (1998), Dow and Endersby (2004) and Schofield (2007). No study to our knowledge has relied on extreme type II distributions (see for example Eaton and Kortum (2002) for an estimation of a model with Fréchet noise in the context of international trade).

the existence of pure strategy equilibria is not always guaranteed in spatial competition models. By contrast, in a stochastic voting theory, pure strategy equilibria typically exist, even in multidimensional and not Euclidian political spaces. The existence of a symmetric equilibrium is demonstrated in the various applications presented in this paper. Second, we consider the entry of purely office motivated candidates, so entry decisions relate solely to the strength of electoral competition, not to policy preferences.

Notable exceptions in the study of multi-candidate elections outside the scope of spatial competition include the works of Besley and Coate (1997), Myerson (1993), Myerson and Weber (1993), Lizzeri and Persico (2001) and Lizzeri and Persico (2005). Besley and Coate (1997) accounts for the effect of strategic voting in a citizen-candidate framework, thereby complementing Osborne and Slivinski (1996), who focus on sincere voting. Myerson (1993) studies electoral equilibria under different electoral rules when candidates simultaneously decide their political platform, so when pure strategy equilibria do not exist. Lizzeri and Persico (2001) and Lizzeri and Persico (2005) apply a similar framework to study the issue of public provision under alternative electoral regimes and the drawbacks of electoral competition. The existence of a symmetric equilibrium is demonstrated in this paper in the various applications of the theory. Furthermore, we suggest that both proportional and run-off systems should lead to policies that are more favorable to the citizenry relative to plurality elections. Relatedly, Lizzeri and Persico (2001) find that public goods are provided less often in winner-take-all system relative to proportional systems.

Coalition formation in the legislature: A large literature has been devoted to the formation of coalitions in the legislature (e.g. Schofield (1993), Baron (1993), Schofield (1997), Baron and Diermeier (2001) and Diermeier, Eraslan, and Merlo (2002)). This literature has studied the duration of coalition in the legislature as well as the convergence of electoral equilibria. We abstract from those issues in this paper and present a version of our theory with endogenous entry that accounts for the formation of coalitions in the legislature in a proportional system. For that purpose, we build on the framework of Baron and Diermeier (2001). Our main contribution is to apprehend the interaction between political platform,

coalition formation and party formation. To our knowledge, coalition politics in the legislature has not dealt with the issue of the formation of political parties and our paper fills this gap. Our main prediction in that respect is that a proportional system makes rent extraction more costly at the margin for existing parties relative to a plurality system. This is why we should expect less parties to form under proportional rules.

Franchise extension in democracy: This work finally relates to the literature on the extension of the franchise and democratization. Lizzeri and Persico (2004) builds a theory of franchise extension and discuss its implications in the case of the nineteenth century Britain. In their model, the extension of the franchise is peaceful and turn politicians away from particularistic politics. This can be valued by a majority within the elites when public good spendings are particularly valuable (relatedly in the theory of Llavador and Oxoby (2005), the extension of the franchise also depends on the existing conflicts within the elite). We also consider a peaceful process of franchise extension. Our central contribution is to uncover that there is a positive income effect that is inherent to the process of franchise extension, independently from the preferences of the citizenry. This explains why it can be consensual for an elite to extend the franchise, given that it can disproportionally benefit from it when the newly enfranchised interests are not politically reactive. Our theory is consistent with Aidt and Jensen (2009a,b, 2013), Acemoglu, Naidu, Restrepo, and Robinson (2015) and Fujiwara (2015) for instance, since they find positive effects of franchise extension on public spendings. Furthermore, we account for an endogenous fragmentation of the polity, although this does not affect our predictions since free entry and office motivations make parties indifferent to the issue of democratization. We also show that the extension of the franchise should be expected if income disparities between the franchised and the disenfranchised are low. This prediction is reminiscent of the studies of Engerman and Sokoloff (2000) and Engerman and Sokoloff (2005) for instance, who provide evidence that greater equality generally lead to the broadening of the franchise. Finally, we suggest that democratization should precede fiscal consolidation, since the income effect of the extension of the franchise is particularly strong when fiscal capacity is weak (the literature on the timing of reforms is still in its infancy, see

for instance Giavazzi and Tabellini (2005)). Finally, our model does not account for the threat of revolution - and does not provide a dynamic analysis of the process of democratization. Arguably, the threat of revolution can be a central determinant of the dynamic process of democratization, as suggested for instance by Acemoglu and Robinson (2000).

3.2 The canonical model

Suppose an individual in the population has a vector of measured attributes s that belongs to a convex and non-empty set S. The vectors $s \in S$ relate to the characteristics of the agents, say their education level, their ethnicity, their wealth, their region of birth and so forth. There are P candidates running for a single-district election indexed $i \in \mathcal{P} = \{1, \ldots, P\}$. We assume a plurality election rule, meaning that the party with the highest vote share wins the election. As only one candidate per party can run in the election, we will use interchangeably the terms party and candidate in the sequel.

We denote q_i the platform of candidate $i \in \mathcal{P}$ in the election, which we assume belongs to a closed and convex set Q_i. A set of feasible policies for candidate is not necessarily an Euclidian space. Components of candidates' platform can relate to tax collection, public good provision, redistribution, alternative institutional arrangements, allocation of natural resource revenues, of campaign resources and so forth. Furthermore, we allow the sets of feasible platforms to be candidate-specific in order to account for factors affecting policies such as differences in candidates' ability or more broadly idiosyncratic constraints that are not directly linked to the election under scrutiny but that weight on candidates' strategic decisions. Political parties for instance can impose such external constraints on their candidates in election.

We assume no commitment issue so the utility from electing candidate i for any individual with attributes s can be written in the form

$$U = V(s, i)\epsilon(s, i) \tag{3.1}$$

where V is non-stochastic. We posit that $V(s, i) \equiv V(s, q_i)$, meaning that the deterministic

component of the utility of the agents with attributes s depends only on the quality of the platform of candidate i, and thus does not reflect some exogenous preference for candidate i. Alternatively, $\epsilon(.,.)$ reflects the idiosyncracies of voting behaviors, which we allow to depend on candidate i rather than on the platform q_i as well as on the attributes $s \in S$. We assume that $\epsilon(s,i)$ is positive for any $s \in S$ and any platform $q_i, i \in \mathcal{P}$.

Any individual votes for the candidate that maximizes his utility. Thus, an agent with attributes $s \in S$ votes for candidate i if

$$V(s,q_i)\epsilon(s,i) > V(s,q_j)\epsilon(s,j) \text{ for any } j \in \mathcal{P} \setminus i. \tag{3.2}$$

The probability that candidate $i \in P$ is chosen by an agent with attribute s is denoted $i(s,q_i,q_{-i})$ and writes:

$$i(s,q_i,q_{-i}) = \Pr[\epsilon(s,j) < V(s,q_i)/V(s,q_j)\epsilon(s,i) \text{ for any } j \in \mathcal{P} \setminus i], \tag{3.3}$$

with q_i the policy of candidate i and q_{-i} the vector of platforms of i's challengers. We denote $F_{(s,i)}$ the cumulative distribution of $\epsilon(s,i)$, which we assume takes support in the set of positive real numbers. The probability $i(s,q_i,q_{-i})$ can be rewritten in the form

$$i(s,q_i,q_{-i}) = \int_0^\infty \prod_{j \in \mathcal{P}\setminus i} F_{(s,j)}(V(s,q_i)/V(s,q_j)\epsilon) dF_{(s,i)}(\epsilon). \tag{3.4}$$

We assume that $\epsilon(i,s)$ is distributed according to Fréchet distributions,

$$F_{i,s}(\epsilon) = \exp(-t_i \epsilon^{-\theta_s}), \tag{3.5}$$

for $\epsilon > 0$, $t_i > 0$ and $\theta_s > 1$ [4], for any $s \in S$ and $i \in \mathcal{P}$. The parameter $t_i > 0$ relates to the concept of valence in the political economy literature. Indeed, this parameter governs the location of the distribution. A higher t_i implies that candidate i on average has a high

4. This restriction is necessary to ensure that the Frechet distribution has a finite mean.

popularity among the citizenry, independently form the platform he promises. The parameter θ_s, which we assume independent from the set of candidates, reflects the amount of variation within the distribution. A higher value of θ_s means that the citizens of type s are highly reactive to the platforms announced by the candidates. We label θ_s the *political responsiveness* of the agents with attributes $s \in S$. From (3.4) and (3.5), we deduce that $vs_i(s, q_i, q_{-i})$ rewrites

$$i(s, q_i, q_{-i}) = \frac{t_i V(s, q_i)^{\theta^s}}{\sum_{j \in \mathcal{P}} t_j V(s, q_j)^{\theta^s}}. \tag{3.6}$$

The vote share of candidate $i \in \mathcal{P}$ is then a function $vs_i : \prod_{j \in \mathcal{P}} Q_j \to [0, 1]$,

$$vs_i(q_i, q_{-i}) = \sum_{s \in S} x_s i(s, q_i, q_{-i}) \tag{3.7}$$

with $x_s \in [0, 1]$ the fraction of agents with attributes $s \in S$ in the population, given that $\sum_{s \in S} x_s = 1$.

Definition 1. *An electoral equilibrium is such that any candidate promises a platform that maximizes his vote share and each candidate expects his challengers to do the same. The equilibrium platform $w_i \in Q_i$ of any candidate $i \in \mathcal{P}$ is such that*

$$w_i = \arg\max_{q_i \in Q_i} vs_i(q_i, w_{-i}) \tag{3.8}$$

for any $i \in \mathcal{P}$ with $w_{-i} \in Q_{-i} = \prod_{j \in \mathcal{P} \setminus i} Q_j$ the vector of optimal platforms of candidate i's challengers.

Assuming that candidates maximize their margin of victory relative to their challengers (i.e. their plurality) will not change the results (see Coughlin (1992) or Coughlin and Nitzan (1981) for such a theory in two-candidate elections). The following theorem is a generalization of two results established in the case of two-candidate elections in random voting models with logit distributions by Coughlin (1992, p. 96-97) (theorem 4.2 and Corollary 4.2).

Theorem 1. *Suppose that there is an election in which (i) there is a finite set of attributes S and $\theta_s > 1$ for any $s \in S$ (ii) there is a finite set of candidates $\mathcal{P}$ and the set of feasible policy*

Q_i of any candidate i is compact and convex and (iii) the voters have $\mathcal{C}^1$ utility functions $V(s,.)$ such that $V(s,.)^{\theta_s}$ is concave over the set of feasible policies $Q = \bigcup_{i \in P} Q_i$. There exists an electoral equilibrium $\{w_i\}_{i \in \mathcal{P}} \in \prod_{i \in \mathcal{P}} Q_i$. The electoral equilibrium is unique if there exists some $s \in S$ such that the function $V(s,q)^{\theta_s}$ is strictly concave on Q [5]

Démonstration. The proof is available in Appendix 3.6.1. □

It has to be noted that this theorem applies outside the scope of redistribution strategies or Euclidian political space. Components of candidates' platform can relate to tax collection, public good provision, redistribution, alternative institutional arrangements, allocation of natural resource revenues, of campaign resources and so forth. As well the set of feasible platforms need not to be the same for all candidates. Note that the theorem requires the functions $V(s,.)^{\theta_s}$ to be quasiconcave (or strictly concave) for an equilibrium to exist (to be unique). This is stronger than the quasi-concavity (strict concavity) of the utility function $V(s,.)$ when $\theta_s > 1$ (ie. the Frechet distributions are assumed to have a well defined finite mean). This will be satisfied when θ_s is not too much larger than 1 compared to the concavity of $V(s,.)$. This is just a general feature of a standard probabilistic voting model to generate a (unique) equilibrium, namely that the distribution of the stochastic element of the utility model to ensure the existence of the equilibrium.

The proof of Theorem 1 follows from the convexity of the maximization problem of candidates, given that the functions $V(s,.)^{\theta_s}$are concave and continuous (the standard existence theorem that is applied is derived from Fudenberg and Tirole (1991, p. 34)). Unicity follows from strict convexity, which is insured whenever at least one function $V(s,.)^{\theta_s}$ is strictly concave.

5. When the function $V(s,q)$ is twice differentiable, it will be the case when the following matrix

$$\left[\frac{\partial^2 V}{\partial q_v \partial q_h}\right] + \frac{(\theta^s - 1)}{V}\left[\frac{\partial V}{\partial q_v}\right] \cdot \left[\frac{\partial V}{\partial q_v}\right]^T$$

defines a semi-defini negative bilinear form.

3.3 Axiomatic approach to probabilistic voting

This section is more technical and readers who wish to skip this may go directly to the next section. As we focus on the agents with attributes $s \in S$, we will denote $i(s, \mathcal{P}) \equiv i(\mathcal{P})$ the probability that an agent of type s votes for candidate i when the set of candidates is $\mathcal{P}$. Similarly, for any subset of candidates $\mathcal{C} \subseteq \mathcal{P}$, we will denote $C(s, \mathcal{P}) \equiv C(\mathcal{P})$ the probability that the candidate chosen by the agents with attributes s belongs to the subset $\mathcal{C}$ of candidates.

Axiom 1. *(Independence of Irrelevant Alternatives Axiom). For all possible alternative set of candidate $C \subseteq P$ and vectors of measured attributes $s \in S$,*

$$i(\mathcal{C})/j(\mathcal{C}) = i(\mathcal{P})/j(\mathcal{P}), \tag{3.9}$$

Axiom 1, adapted from McFadden (1974), says that the odds of i being chosen relative to candidate j out of $\mathcal{P}$ candidates $i(s, \mathcal{P})/j(s, \mathcal{P})$ are equal to the odds of choosing i relative to j out of any subset of candidates $\mathcal{C}$, $i(s, \mathcal{C})/j(s, \mathcal{C})$.

Axiom 2. *(Positivity). For all possible alternative set of candidates $\mathcal{C} \subseteq \mathcal{P}$ and vectors of measured attributes $s \in S$, $i(s, \mathcal{C}) > 0$.*

For any set of candidates $\mathcal{C}$, an agent with attributes $s \in S$ has a strictly positive probability of voting for *any* candidate $i \in \mathcal{C}$. The main consequence of Axiom 2 is that deterministic voting models are not consistent with an axiomatic approach.

We now define a special class of probabilistic voting models originally introduced by Luce (1959), which will be particularly useful in the subsequent analysis. Using the framework introduced by Becker, Degroot, and Marschak (1963), a probabilistic voting model will be called a *Luce or strict voting utility model* if there exist positive "utility indicator" functions $v(s, i)$ for any candidate $i \in \mathcal{C} \subseteq \mathcal{P}$ such that the probability of i being chosen out of $\mathcal{C}$ by

the agents with attributes s can be expressed as

$$i(s,\mathcal{C}) = \frac{v(s,i)}{\sum_{j\in\mathcal{C}} v(s,j)}. \tag{3.10}$$

Theorem 2. *A probabilistic model satisfies axioms 1 and 2 if and only if it is a Luce voting model.*

Instead of giving a full proof of the preceding theorem - which can be found in Becker, Degroot, and Marschak (1963) (their Theorem II) - we follow the ingenuous method of McFadden (1974) in order to characterize $j(\mathcal{P})$. Assume that there are only two candidates $\{i,j\}$ competing in the election. From (3.9) with $\mathcal{C} = \{i,j\}$,

$$i(\mathcal{P}) = \frac{i(\{i,j\})}{j(\{i,j\})} j(\mathcal{P}). \tag{3.11}$$

Thus,

$$\frac{i(\{i,j\})}{j(\{i,j\})} = \frac{i(\mathcal{P})/k(\mathcal{P})}{j(\mathcal{P})/k(\mathcal{P})} \tag{3.12}$$

for some third candidate $k \in \mathcal{P} \setminus \{i,j\}$, implying that

$$\frac{i(\{i,j\})}{j(\{i,j\})} = \frac{i(\{i,k\})/k(\{i,k\})}{j(\{j,k\})/k(\{j,k\})}. \tag{3.13}$$

Moreover,

$$\sum_{i\in\mathcal{P}} i(\mathcal{P}) = 1 = \sum_{i\in\mathcal{P}} \frac{i(\{i,j\})}{j(\{i,j\})} j(\mathcal{P}). \tag{3.14}$$

Thus,

$$j(P) = \frac{1}{\sum_{i\in P} i(\{i,j\})/j(\{i,j\})}. \tag{3.15}$$

From (3.12), we deduce that

$$j(\mathcal{P}) = \frac{j(\{j,k\})/k(\{j,k\})}{\sum_{i\in P} i(\{i,k\})/k(\{i,k\})}. \tag{3.16}$$

We denote $j(s,\{j,k\})/k(s,\{j,k\}) = v(j,k,s)$ so that (3.12) rewrites

$$j(\mathcal{P}) = \frac{v(j,k,s)}{\sum_{i\in\mathcal{P}} v(i,k,s)}, \tag{3.17}$$

meaning that the probability of an agent with attribute s from voting j out of $\mathcal{P}$ candidates is equal to how well j fares against some candidate k in a pairwise election relative to how well all the candidates fare against the same candidate k in pairwise elections. Our last axiom gives a specification for $v(j,k,s)$, which is the main departure from McFadden (1974).

Axiom 3. *(Irrelevance of Alternative Set). The function $v(j,k,s)$ determining the selection probability in pairwise elections has a product separable form and there exist utility indicators $u(q_j,s)$ and $u(q_k,s)$ and some constants $t_j > 0$, $t_k > 0$ and $\theta_s > 0$ for any pair of candidates $j,k \in \mathcal{P}$ and any attributes $s \in S$ such that*

$$v(j,k,s) = \frac{t_j u(q_j,s)^{\theta_s}}{t_k u(q_k,s)^{\theta_s}}. \tag{3.18}$$

We posit in Axiom 3 that in pairwise elections, there are three dimensions in individual voting decisions. The first dimension of voting decisions is linked to the *popularity* or valence of the candidates. The parameters t_i and t_k model the popularity of the two candidates and are assumed independent from individual attributes $s \in S$ and from the promised platforms q_j and q_k.

The second aspect of voting decisions that is accounted for in (3.18) is linked to the *quality* of the promised platforms, as evaluated by the agents with attributes $s \in S$. This dimension of voting decisions is dealt with through utility indicator functions $u(.,s)$ that are defined over the set of feasible platforms $Q = \bigcup_{i\in\mathcal{P}} Q_i$ for any vector of attributes $s \in S$.

The last parameter θ_s is linked to the rate of substitution between popularity and quality. If $u(q_j,s)/u(q_k,s)$ - the relative quality of the platform of candidate j - increase by a percentage point, then the relative popularity of candidate j must decrease by θ_s percentage point so that the odds of electing j stay constant. Thus, when θ_s is high, even popular candidates

need to provide quality policies because popularity does not affect much voting decisions.

Axiom 3 sets a specification for the probability of voting any candidate j relative to some challenger k in pairwise elections. To illustrate, McFadden (1974) assumes an additively separable form in his Axiom of Irrelevance of Alternative Set. The author then establishes that the only distribution of the noise parameters affecting individual decisions that are compatible with his three behavioral axioms are Weibull distributions. With the product functions of (3.18), we will show that the noise in voting decisions must be distributed according to Fréchet distributions, which are as well labeled *inversed* Weibull functions.

Lemma 1. *A probabilistic voting model satisfies the three preceding Axioms if and only if for any vector of attributes $s \in S$ and for any candidate $j \in \mathcal{P}$ there exists a utility indicator $u(.,s)$ defined over the set of feasible policies Q, a popularity parameter $t_j > 0$ and a political responsiveness parameter θ_s such that*

$$j(\mathcal{P}) = \frac{t_j u(q_j, s)^{\theta_s}}{\sum_{i \in \mathcal{P}} t_i u(q_i, s)^{\theta_s}}. \tag{3.19}$$

This result is obtained by applying theorem 2 and substituting (3.18) in (3.16). We will next derive the family of distributions that satisfy the behavioral assumptions given in the three preceding Axioms. Summarizing our previous findings, from (3.4), we know that a random utility voting model with a specification given in (3.1) is such that the probability for an individual s to vote for candidate i out of $\mathcal{P}$ is

$$i(s, \mathcal{P}) = \int_0^{\infty} \prod_{j \in \mathcal{P} \setminus i} F_{(s,j)}(V(s,i)/V(s,j)\epsilon) dF_{(s,i)}(\epsilon)$$

where $F_{(s,i)}(.)$ is the distribution of the random parameter $\epsilon(s,i)$.

From the preceding formal development inspired of McFadden (1974), we know that a probabilistic voting model respecting Axioms 1, 2 and 3 must be such that for any vector of attributes $s \in S$ and for any candidate $i \in \mathcal{P}$, the probability of an individual s to vote for

candidate i out of $\mathcal{P}$ takes the form

$$j(s,P) = \frac{t_j u(q_j,s)^{\theta_s}}{\sum_{i\in P} t_i u(q_i,s)^{\theta_s}}.$$

with $u(q_k,s)$, $k \in \mathcal{P}$ some utility indicators, $t_k > 0$ a popularity indicator and θ_s the political responsiveness of the agents with attributes $s \in S$.

We define the following class of distributions.

Definition 2. *Let $\mathcal{C}_d$ the class of distributions such that for any pair of candidates $i,j \in \mathcal{P}$ there exists a strictly positive constant $\alpha_{i,j}$ such that $F_{i,s}(\epsilon) = F_{j,s}(\alpha_{i,j}\epsilon)$.*

This class of distributions includes more specifically the case where the noise parameters are all distributed according to the same distribution (when $\alpha_{i,j} = 1$ for any pair $i,j \in \mathcal{P}$) and allows for some differences between the distributions which are a matter of translation.

Theorem 3. *In the class of distribution $\mathcal{C}_d$, a probabilistic voting model satisfies axioms 1, 2 and 3 if and only if it is a random utility model where the noise parameters $\epsilon(i,s)$ are distributed according to Fréchet distributions $F_{i,s}(\epsilon) = exp(-t_i\epsilon^{-\theta_s})$ for some positive parameters t_i and θ_s.*

Démonstration. The proof is available in Appendix 3.6.2. □

This concludes the micro-foundations of random voting utility models with random parameters distributed according to some Fréchet (or extreme type II) distributions. As in the approach to economic decisions of McFadden (1974), we have demonstrated that voting probabilities can be interpreted as deriving from representative utilities, which are affected by the popularity of the candidates and by the quality of the political platforms they offer.

Furthermore, this theory makes it simple to ascertain the effect of an increased number of candidates in election on voting decisions and therefore provides an approach to the industrial organization of politics. From the Luce model implied by the two first axioms, it nevertheless inherently builds in the model a particular effect of competition on voting behavior. Indeed, an increase in the set of competing parties necessarily leads to proportional decreases in

the vote shares of the old candidates, and a corresponding increase of the vote share of new candidates. This is the first main limitation of this theory.

The second main limitation of our approach lies in the specification of selection probabilities in pairwise election given in Axiom 3. Indeed, although we are able to specify probabilistic voting behaviors in pairwise elections, our approach assumes that voting decisions respond to three main dimensions that have found support in the political economy literature. The first one is the valence of candidates in election (the parameters that we denoted t_i, $i \in \mathcal{P}$). The second is the quality of the political platforms that are proposed by the candidates in election. We modeled this by assuming that the agents of type $s \in S$ derive some utility $u(q, s)$ from electing a candidate that implements some platform q. Finally, we have assumed that the last dimension of probabilistic voting decisions is linked to the elasticity of substitution between political platforms' quality and the valence of candidates in election through the parameters θ_s, $s \in S$. More complex approaches may account for other dimensions of probabilistic voting decisions and will find different distributions for the randomness in voting decisions. Our approach is however sufficiently simple and flexible to be adapted to different topics of the political economy literature as we demonstrate in the next section.

3.4 Applications

Our aim in this section is to demonstrate that our analytical framework provides a unifying approach for various topics of the political economy literature. The two first subsections study special interest politics. In the third application, we extend the framework to a richer political space by accounting for public provision and special interest spendings. In the fourth section, we compare the endogenous entry of parties under the runoff and the plurality systems. The fifth application accounts for the formation of coalitions in the legislature under proportional rules while the sixth and final application studies the extension of the franchise in democracy when the fragmentation of the polity is endogenous.

3.4.1 Application 1: Special interest politics with heterogeneous voters and homogeneous entrants

We provide in this application a simple theory of special interest politics that accounts for the endogenous formation of political parties. We assume in this application that the set of individual attributes $s \in S$ characterize a set of interest groups. Let $S = \{1, \ldots, N\}$. The size of group s is denoted n_s, n gives the size of the economy and $x_s = n_s/n$ the fraction of agents of type s. We denote P^e the expected number of parties and posit that the political entrepreneurs that will create the parties have rational expectations. Since $P^e \equiv P$ in equilibrium, we will abuse the notations by denoting P the expected number of parties in the sequel. We will also use the notation $\mathcal{P} \equiv P$ when it is not confusing. Once formed, parties compete for the votes of the citizens in a single district plurality election.

We assume that the utility of the agents in the interest group $s \in S$ only depends on their consumption that we denote c_s. Let y be the aggregate income, $y = \sum_{s \in S} n_s y_s$ We denote $\overline{\theta} = \sum_{s \in S} x_s \theta_s$ and $\overline{y} = y/n$ the average political responsiveness and income respectively. We assume the following specification for the utility function of the agents in group s,

$$u(c_s) = c_s^{1-\epsilon}/(1-\epsilon) \tag{3.20}$$

with $0 < \epsilon < 1$ a parameter that captures the degree of diminishing returns to private consumption. Indeed, if ϵ is small, the marginal utility of consumption falls slowly as the level of consumption rises. Thus, even at high consumption levels, the utility is significantly affected by variations in consumption when ϵ is small. By contrast when the value of ϵ is high, the utility levels of richer groups are less affected by variations in consumption levels.

Party $i \in P$ transfers $\tau_{s,i}$ units of income to any any agent with attributes $s \in S$. We assume that the preceding transfer can take negative values although it is bounded by the initial wealth of group s (i.e. $\tau_{s,i} \geq -y_s$ necessarily). Thus, when party $i \in P$ is in office, the consumption of any agent in group s is equal to the sum of his revenues plus the

group/party-specific transfers,

$$c_s(q_i) = y_s + \tau_{s,i} \tag{3.21}$$

We assume that the transfers do not sum to zero, as party $i \in P$ extracts a fraction $\chi^i \in [0,1]$ of the tax base as rents,

$$\sum_{s \in S} n_s \tau_{s,i} = -\chi_i y. \tag{3.22}$$

In those settings, the policy vector of party i is $q_i = \{\chi_i, \{\tau_{s,i}\}_{s \in S}\}$ and belongs to a convex set Q_i since the feasible transfers are bounded ($\tau_{s,i} \in [-y_s, \sum_{p \neq s} n_p y_p / n_s]$ where the higher bound corresponds to a policy that transfers all the resources to group s while no rents are extracted from the citizenry). Observe additionally that the aggregate consumption is such that

$$\sum_{s \in S} n_s c_s(q_i) = y + \sum_{s \in S} n_s \tau_{s,i} = y(1 - \chi_i). \tag{3.23}$$

Applying the canonical model of section 3.2, an agent with attributes $s \in S$ prefers candidate i if

$$u(c_s(q_i))\epsilon(s,i) > u(c_s(q_j))\epsilon(s,j) \text{ for any } j \in P \setminus i, \tag{3.24}$$

given that $\epsilon(s,i)$ follows some Fréchet distribution $F_{i,s}(\epsilon) = \exp(-t_i \epsilon^{-\theta_s})$ for any $s \in S$ and any $i \in P$.

To allow for some aggregate uncertainty in terms of the outcome of the political competition, we assume that political preferences does not directly translate into votes. More specifically, we posit that the probability that an agent with attributes s votes for candidate i writes:

$$P_i(s, q_i, q_{-i}) = \eta(s,i,P)\Pr[\epsilon(s,j) < u(c_s(q_i))/u(c_s(q_j))\epsilon(s,i) \text{ for any } j \in P \setminus i], \tag{3.25}$$

with $\eta(s,i,P) \in [0,1]$ the probability that an agent with attributes s that prefers candidate i in the set P goes to the ballot. This additional parameter translates the idea that the probability that an agent goes to the ballot depends both on his characteristics (i.e. on s)

and on the candidate he is expecting to support. Indeed, there is always some randomness associated to each candidate on how they can effectively mobilize people that support them politically to go to the ballot box (this may reflect the discrepancy between opinion surveys and actual vote behavior).

As a matter of simplification, we assume that $\eta(s, i, P)$ is independent of $s \in S$, meaning that any agent that intends to vote for candidate $i \in P$ has the same likelihood of casting a ballot. We use the notation $\eta(s, i, P) \equiv \eta(i, P)$ hereafter. Using the formalization of the preceding section, we can rewrite the probability that an agent with attributes s votes for candidate i as

$$P_i(s, q_i, q_{-i}) = \eta(i, P) \frac{t_i u(c_s(q_i))^{\theta_s}}{t_i u(c_s(q_i))^{\theta_s} + \sum_{j \in P \setminus i} t_j u(c_s(q_j)),^{\theta_s}} \tag{3.26}$$

and using the law of large number, we can deduce that the vote share of candidate i takes the form

$$vs_i(q_i, q_{-i}) = vs_i^t(q_i, q_{-i}) \eta(i, P), \tag{3.27}$$

with

$$vs_i^t(q_i, q_{-i}) = \sum_{s \in S} \frac{n_s}{n} \frac{t_i u(c_s(q_i))^{\theta_s}}{t_i u(c_s(q_i))^{\theta_s} + \sum_{j \in P \setminus i} t_j u(c_s(q_j))^{\theta_s}}. \tag{3.28}$$

As per (27), the vote share of candidate i is the product of a "theoretical" vote share vs_i^t, which is the fraction of people that prefer candidate i in the population, with the probability $\eta(i, P)$ that those people go to the ballot box. We posit that the probability $\eta(i, P)$ takes the following form:

$$\eta(i, P) = \frac{\mu_i}{\sum_{k \in P} \mu_k}, \tag{3.29}$$

with μ_j a warm glow for casting a ballot for candidate i that we assume distributed according to a Fréchet distribution $G(\mu) = exp(-\mu^{-K})$ with $K > 1$. Given this, one may express the probability for a candidate i to get the largest fraction of effective votes as

$$G_i(q_i, q_{-i}) = \Pr[\mu_j < vs_i^t(q_i, q_{-i}) / vs_j^t(q_i, q_{-i}) \mu_i) \text{ for any } j \in \mathcal{P} \setminus i],$$

$$\begin{aligned} G_i(q_i, q_{-i}) &= \int_0^{\infty} \left[G((vs_i(q_i, q_{-i})/vs_j(q_i, q_{-i})\eta)\right]^{P-1} dG(\mu). \\ &= \frac{\left[vs_i^t(q_i, q_{-i})\right]^K}{\sum_{j\in\mathcal{P}} \left[vs_j^t(q_i, q_{-i})\right]^K}. \end{aligned}$$

An electoral equilibrium platform $(q_i^*)_{i\in P} = \{\chi_i^*, \{\tau_{s,i}^*\}_{s\in S}\}_{i\in P}$ is a set of platform such that each party seeks to maximize its expected rents given what the other parties propose to the voters. So

$$q_i^* = \{\chi_i^*, \{\tau_{s,i}^*\}_{s\in S}\} = \underset{q_i=\{\chi_i,\{\tau_{s,i}\}_{s\in S}\}}{\arg\max} \; y\chi_i \cdot G_i(q_i, q_{-i}^*), \tag{3.30}$$

with q_{-i}^* the vector of optimal platforms of i's challengers.

We assume homogeneous candidates in the election, meaning that $t_i = t_j$ for any $i, j \in P$. Hence

$$vs_i^t(q_i, q_{-i}) = \sum_{s\in S} \frac{n_s}{n} \frac{u(c_s(q_i))^{\theta_s}}{u(c_s(q_i))^{\theta_s} + \sum_{j\in P\setminus i} u(c_s(q_j))^{\theta_s}}$$

and we characterize a symmetric equilibrium in policies q_i. The first-order condition with respect to the transfers to group s writes

$$-n_s G_i(q_i, q_{-i}^*) + y\chi_i \left[\frac{\partial G_i}{\partial vs_i}\frac{\partial vs_i}{\partial \tau_{s,i}} + \sum_{j\neq i} \frac{\partial G_i}{\partial vs_j}\frac{\partial vs_j}{\partial \tau_{s,i}}\right] = 0 \tag{3.31}$$

Condition (3.31) says that when party i marginally increases the transfers to group s, on the one hand, it reduces the expected level of extraction of an amount $n_s G_i(q_i, q_{-i}^*)$. On the other hand, party i grabs an increased vote share from his challengers (the second term in bracket) which comes from an increase in his own vote share vs_i and a decrease in the vote shares of all of his the challengers for a given interest group s. We show the following result in the Appendix.

Proposition 1. *A symmetric equilibrium necessarily exists when parties are homogeneous and is such that for any party i, and any $k, p \in S$*

$$\frac{\theta_s}{c_s(q_i)} = \frac{\theta_p}{c_p(q_i)} = \frac{1}{\frac{y}{n}\chi_i K(1-\epsilon)(1-G_i)}. \tag{3.32}$$

and the consumption of the agents in group s as a function of the rents extracted is

$$c_s(q_i) = \frac{\theta_s}{\overline{\theta}}\overline{y}(1-\chi_i). \tag{3.33}$$

In equilibrium, each group $s \in S$ gets a fraction $\theta_s/\overline{\theta}$ of the transferable revenues. This result derives directly from the fact that - notwithstanding the effect of the responsiveness parameters $\{\theta_s\}_{s\in S}$ - there always is a higher marginal benefit at targeting the poorest groups precisely because they have lower consumption levels.[6] The main consequence of this is that the optimal redistribution scheme can be understood as proceeding in two steps. First, it consists in neutralizing the effect of the income distribution on the vote share by means of transfers. Then, it consists in redistributing resources according to the distribution of the responsiveness parameters across income groups. Ex-post and ex-ante income inequalities are consequently independent since the latter reflects the political behaviors of the existing interest groups. This result obtains because political institutions are built independently from the income distribution. Indeed, parties compete in democratic elections and parties and interest groups have no vested interests so the former exploit "efficiently" the distribution of income so as to maximize their level of extraction. Doing so simply shifts the distribution of income toward the most politically reactive groups.[7]

Given that the electoral equilibrium is convergent, by substituting (3.33) in the first-order condition (31), we find that the rents χ_i are such that so

$$\chi_i = \frac{1}{1+\overline{\theta}K(1-\epsilon)(P-1)/P}. \tag{3.34}$$

The rents χ_i decrease with the number of parties competing in the election. Indeed, the marginal loss of the parties from extracting more rents out of the citizenry is higher when

6. This effect has been noticed already by Dixit and Londregan (1996).

7. Note that the preceding result implies that the absence of vested interests between parties and interest groups does not necessarily imply that resources are to be transferred from high income to low income groups in equilibrium. Indeed, if the rich are significantly more politically responsive than the poor, then we should expect the contrary. Thus, high levels of income inequalities could directly relate to a differential in the ability to coordinate voting behavior across interest groups.

each party has more competitors.

Given that the rents extracted by any party decrease with P, the expected utility derived by any party from winning the election is a decreasing function of the number of existing parties P. In order to derive a closed form result for the number of parties competing in this election, we assume a fixed cost of party formation c. The equilibrium number of parties P^* is then uniquely determined and is the highest integer below the solution of

$$\frac{1}{P}\chi_i y = c, \tag{3.35}$$

with χ_i given in (3.34).

Proposition 2. *The number of parties competing in the election is the highest integer below* P^* *with*

$$P^* = \frac{1}{1 + K\overline{\theta}(1-\epsilon)}[K\overline{\theta}(1-\epsilon) + \frac{y}{c}] \tag{3.36}$$

if $y > c$*, and* $P^* = 0$ *otherwise.*

- P^* *decreases with the cost of party formation* c*, with the average responsiveness of the citizenry* $\overline{\theta}$ *and the shape of the aggregate uncertainty* K*. Alternatively,* P^* *increases with the income of the citizenry* y*, and with the degree of diminishing returns to private consumption* ϵ*.*
- *An increase in the size of any group* n_s *that has a lower than average responsiveness* $\theta_s < \overline{\theta}$ *increases the number of parties while the effect of an increase in the size of a group with a higher than average responsiveness is ambiguous.*

Démonstration. The proof is available in Appendix 3.6.4. □

Note first that if $y < c$, there is no party formation because the entry cost c is higher than the maximum level of rents that can be extracted by any party from the citizenry. We assume that the condition $y < c$ is respected in the sequel.

Relative to the comparative statics, a higher value of the cost of party formation straightforwardly affects negatively the fraction of parties P^*. Additionally, a higher $\overline{\theta}$ increases

the marginal cost at capturing rents, since it implies that voters are more responsive on average to political platforms in their voting behavior. By the same token, an increase in the responsiveness of any group θ_s affects negatively the number of parties in equilibrium. Similarly, whenever the degree of diminishing returns to private consumption ϵ increases, then the agents care less about higher consumption levels and the marginal cost at increasing rents decreases for the existing parties. This, in turn, affects positively the number of parties willing to enter the political arena in equilibrium. Alternatively, if the average income increases, then so does the amount of rents that can be extracted and the party fragmentation increases.

Regarding the last point of the proposition, observe that an increase in the size of a group n_s affects both the tax base and the average responsiveness. Indeed, on the one hand, when n_s increases, so does the tax base and this tends to increase P^*. On the other hand, an increase in n_s also affects the average responsiveness $\overline{\theta}$, which will increase when $\theta_s > \overline{\theta}$ and decrease otherwise. Thus, when the size of a group that has a lower than average responsiveness increases, then the number of parties in equilibrium necessarily increases because (i) the taxable income increases and (ii) the average responsiveness decreases, so the marginal cost at capturing rents decreases. Alternatively, when $\theta_s > \overline{\theta}$, then if the taxable income still increases, the average responsiveness decreases so the overall effect of n_s on P^* is ambiguous.

We close this first application by substituting P^* in (3.34) in order to find the gross amount of rents η^* captured per party and the equilibrium consumption levels,

$$\eta^* = y\chi_i = \frac{1}{1 + (1-\epsilon)K\overline{\theta}}((1-\epsilon)K\overline{\theta}c + y), \tag{3.37}$$

and

$$c_s^* = \frac{K\theta_s(1-\epsilon)}{n(1 + K\overline{\theta}(1-\epsilon))}(y - c) \tag{3.38}$$

for any $s \in S$. The comparative statics are summarized in the next proposition.

Proposition 3.

— *A higher responsiveness θ_s increases the consumption of the agents in group s while it decreases both the rents η^* and the consumption of any other group $p \neq s$. Alternatively,*

a higher degree of decreasing returns to consumption ϵ increases the rents η^* and decreases the consumption of any group $s \in S$. Similarly, an increase in the aggregate uncertainty K increases the rents η^* and decreases the consumption of any group $s \in S$. Alternatively, a higher income level y_s will affect positively both the rents η^* and the consumption of any group $p \in S$. Finally, we should expect a higher cost of party formation c to increase the level of extraction η^* and to decrease the consumption levels c_s^* for any $s \in S$.

— *If n_s increases, so do the rents if group s has a lower than average responsiveness θ_s. The effect is ambiguous if group s has a higher than average responsiveness. Similarly, an increase in n_s affects positively the consumption level of any group $p \in S$ (including s) when group s has a sufficiently low reactivity or a sufficiently high income y_s.*

Whenever the responsiveness of any group s increases, it becomes more costly in terms of vote share to capture rents, so the parties optimally reduce them. Moreover, the marginal benefit at redistributing resources to group s increases, while the marginal cost at distributing resources to any other group decreases. It is then not surprising that c_s^* increases with θ_s while c_p^* decreases for any $p \neq s$. This result accords with the standard prediction of theories where elections determine the allocation of resources that increased political participation of some interest groups leads to more redistribution toward them (e.g. Lindbeck and Weibull (1987), Dixit and Londregan (1996)).

Furthermore, we can reasonably assume that more informed voters should cast their vote on candidates' platforms rather than on non-policy relevant matters (what we have labeled popularity or valence in section 3.2). Thus, we should expect more informed voters to be more responsive in their voting behavior, where responsiveness is the elasticity of substitution between quality and popularity (see our axiomatic approach to probabilistic voting in section 3.3). The preceding result is then consistent with series of evidence in the literature on political economy and media that show that better informed voters get higher transfers. [8] If

8. Strömberg (2004) finds for example that the expansion of radio in the United States in the early 1920s created rapid and substantial changes in economic policy making. He finds that a one-standard deviation increase in the share of households with radios in a certain county led governors to increase per capital relief

the empirical literature has been concerned with linkage between voters' information and public policies, it has not studied yet the effect of information on party fragmentation but as per Proposition 2, we should expect a low average responsiveness of the citizenry to increase the party fragmentation.

According to Proposition 3, when the degree of decreasing returns to private consumption ϵ increases, the marginal utility of consumption is lowered so parties are incentivized to capture more rents. This translates in equilibrium in lower transfers to the citizenry.[9] This result is not surprising since ϵ enters the model exactly as the responsiveness parameters.

According to Proposition 3, when the income level of any group increases, so does the tax base. Thus, both the rents captured by the parties and the consumption levels of the voters increase in equilibrium. Richer countries should have more fragmented polities given that the other parameters are held constant.

It is clear that the cost of party formation c has a positive effect on rent capture. Indeed, when the cost of party formation is higher, then parties are willing to enter the political arena only if it allows them to capture high rents. In equilibrium, this translates in a lower degree of competition on the political market, and consequently implies lower transfers to the citizenry. Observe that parties have an interest in deterring further entrance in the political arena by raising c, while the citizenry would prefer low entry costs so as to decrease the extent of rent capture by making the political market more competitive. As a simple illustration of the preceding result, Scherlis (2014) argues that a legitimacy crisis and popular discontent led to important reforms aiming at reducing party formation costs across Latin American countries in the 1990s. The new Colombian Constitution enacted in 1991 virtually abolished entry barriers to democratic competition by recognizing social movements and other groups of citizens as equivalents to parties (Scherlis (2014)). As of 1994, 50,000 signatures or 50,000 votes

spending by 9 percent and implied as well increasing transfers to rural voters. Similarly, Besley and Burgess (2002) show that Indian state governments are more responsive to falls in food production and crop flood damage via public good distribution and calamity relief expenditure where newspaper circulation is higher and electoral accountability greater. See as well Snyder and Strömberg (2010) for similar results and the review of Prat and Strömberg (2011).

9. This result is fully consistent with that previously established by Dixit and Londregan (1996) with the same utility specification (3.20) for the voters.

in the preceding elections were required to obtain legal recognition. Furthermore, anyone could register a candidacy, even without legal party recognition, by paying a sum to be refunded if a threshold of 50,000 votes is reached. Finally, multiple lists from the same party were authorized as well. As a result, the number of lists competing for the Senate and the House of Representatives steadily increased from 1990 to 2002. Interestingly, once policymakers and parties regained legitimacy, new laws aiming at reinstating high party formation costs where voted. Indeed, President Uribe, supported by conservatives, liberals, and a part of the leftist Democratic Pole reformed the party system so as to reduce the fragmentation of the polity. The threshold for legal recognition was increased to 2% of the votes for example. Similar reversal of the reforms on the costs of party formation are observed in periods of economic growth in Argentina, Mexico and Peru (Scherlis (2014)).

Relative to the second point of the proposition, note that an increase in the size of group s can affect political strategies through two channels, the average responsiveness $\overline{\theta}$ and the tax base y. Whenever the size of a group with a lower than average responsiveness increases, the marginal cost at increasing rents decreases necessarily so the parties can capture more rents. This is because a higher size of a group with a low reactivity (i) decreases the average responsiveness of the citizenry and (ii) increases the taxable revenues. Alternatively, when the size of a group with a higher than average responsiveness increases, the effect on rents is ambiguous since on the one hand the tax base increases, while on the other hand the higher average responsiveness puts a downward pressure on rent extraction.

The effect of a higher n_s on the consumption levels is relatively similar. Indeed, on the one hand, a higher value of n_s will affect the average responsiveness. On the other hand, a higher value of n_s will increase the tax base and thus increase the transfers received by the interest groups (this is a standard income effect). Thus, if the size of a group with a sufficiently low reactivity increases, we should expect the reactivity effect to dominate, and thus the transfers to decrease. At the opposite when the agents in group s have a sufficiently high income y_s, then the income effect dominates. [10]

10. It can be shown by deriving (42) that c_s increases with n_s whenever $y_s > (n\overline{y} - c)(1 + \theta_s)/(n(1 + \overline{\theta}))$.

3.4.2 Application 2: Special interest politics with homogeneous voters and heterogeneous entrants

In order to extend the preceding framework to the case of heterogeneous political parties, we will proceed in two steps. First, we will consider a problem of sequential entry where the popularity of the candidates running for the election is exogenously given. Second, we will apprehend the case of simultaneous entry with endogenous popularity.

Sequential entry in a simple case

Assume that there is initially one party competing in the election. This party has a high popularity that we denote t_h. This may be because it benefits from significantly larger fundings than its challengers. We do not assume that this party has the ability to deter the entry of potential challengers. Parties with a lower popularity can decide to run for the election. We denote t_l the popularity of the potential entrants, with $t_l < t_h$. Since we assume homogeneous voters, the platform of any party i reduces to a level of extraction χ_i. The consumption of any citizen when i wins the office simplifies to

$$c(\chi_i) = y(1 - \chi_i). \tag{3.39}$$

Following the steps of the preceding section, we can easily show that the probability of party $i \in \{h, l\}$ being elected now writes as

$$\begin{aligned} G_h(\chi_h, \chi_l) &= \frac{[vs_h(\chi_h, \chi_l)]^K}{[vs_h(\chi_h, \chi_l)]^K + (P-1)\,[vs_l(\chi_l, \chi_h)]^K} \\ G_l(\chi_h, \chi_l) &= \frac{[vs_l(\chi_h, \chi_l)]^K}{[vs_h(\chi_h, \chi_l)]^K + (P-1)\,[vs_l(\chi_l, \chi_h)]^K} \end{aligned}$$

with

$$\begin{aligned} vs_h(\chi_h,\chi_l) &= \frac{t_h u(c(\chi_h))^\theta}{t_h u(c(\chi_h))^\theta + (P-1)t_l u(c(\chi_l))^\theta} \\ vs_l(\chi_l,\chi_h) &= \frac{t_l u(c(\chi_l))^\theta}{t_h u(c(\chi_h))^\theta + (P-1)t_l u(c(\chi_l))^\theta} \\ \frac{vs_l(\chi_l,\chi_h)}{vs_h(\chi_h,\chi_l)} &= \frac{t_l u(c(\chi_l))^\theta}{t_h u(c(\chi_h))^\theta}. \end{aligned}$$

Party i still seeks to maximize its expected rents, so

$$\chi_i^* = \arg\max_{\chi_i} \chi_i y G_i(\chi_i, \chi_{-i}^*). \tag{3.40}$$

Following the steps of the previous section, it can be shown that the first-order conditions associated with the preceding optimizations are

$$1 - \frac{1}{n}\frac{\chi_h}{(1-\chi_h)} K(1-\epsilon)\theta \left(1 - \frac{t_h^K}{t_h^K + (P-1)\, t_l^K Z(\chi_h,\chi_l)^K}\right) = 0 \tag{3.41}$$

for the high popularity party, and

$$1 - \frac{1}{n}\frac{\chi_l}{(1-\chi_l)} K(1-\epsilon)\theta \left(1 - \frac{t_l^K}{(P-1)\, t_l^K + t_h^K / Z(\chi_h,\chi_l)}\right) = 0 \tag{3.42}$$

for his $P-1$ challengers, with

$$Z(\chi_h,\chi_l) = \left(\frac{1-\chi_l}{1-\chi_h}\right)^{(1-\epsilon)\theta K}. \tag{3.43}$$

From this point, (3.41) gives a locus $\chi_h(\chi_l, t_l, t_h, P)$, while (3.42) provides a locus $\chi_l(\chi_h, t_l, t_h, P)$. Notice that the unicity of the electoral equilibrium is not straightforward, since there is a complementarity between rent extraction of the two types of parties in equilibrium. Indeed, if the high popularity party captures more rents, then it becomes less costly for low popularity parties to do the same and reciprocally. However, by combining the two first-order conditions, we can establish a third and simple relationship between χ_l and χ_h that proves the unicity

of the electoral equilibrium. Assuming that the cost of party formation is c, the free entry condition still writes as

$$G_l(\chi_h, \chi_l)\chi_l y = c.$$

Proposition 4.

— *There exists a unique electoral equilibrium where the parties with popularity t_l (resp. t_h) extracts a level of rents $\chi_l^*(t_l, t_h, P)$ (resp. $\chi_h^*(t_l, t_h, P)$), with $\chi_l^*(t_l, t_h, P) < \chi_h^*(t_l, t_h, P)$.*

— *$\chi_l^*(t_l, t_h, P)$ increases with t_l and decreases with t_h, while $\chi_h^*(t_l, t_h, P)$ decreases with t_l and increases with t_h. Consequently, $\chi_l^*(t_l, t_h, P) < \chi^s(P) < \chi_h^*(t_l, t_h, P)$ for a given value of P, with $\chi^s(P)$ the level of extraction in the equilibrium where the candidates are homogeneous.*

— *Since $\chi_l^*(t_l, t_h, P)$ decreases with P, the number of entrants is uniquely determined and less parties should form in equilibrium relative to the case where candidates are homogeneous.*

Démonstration. The proof is available in Appendix 3.6.5 □

The intuition of this proposition is represented in Figure 3.1. The idea is that by combining (3.41) and (3.42), we can establish a relatively simple relation between χ_h and χ_l in equilibrium that is represented by the blue curve $\chi_l(\chi_h)$ in figure 3.1. Since this relation is independent from t_h and t_l, the comparative statics are easily established. Of particular interest, the complementarity between χ_l^* and χ_h^* along the equilibrium path is always dominated. For instance, an increase in t_h will lead to higher rents captured by the popular party and lower rents captured by its challengers. This implies that as long as $t_h \neq t_l$, we should expect an electoral equilibrium where the platforms of the two types of candidates are necessarily different. This result holds even if the difference in the valence of the two types of candidates is small, by contrast with the mean voter theorem of Schofield (2007).

Furthermore, it is easy to show that lower popularity parties will reach strictly lower vote shares relative to the case where candidates are homogeneous ($\chi_l^* < \chi^s$, as represented on the

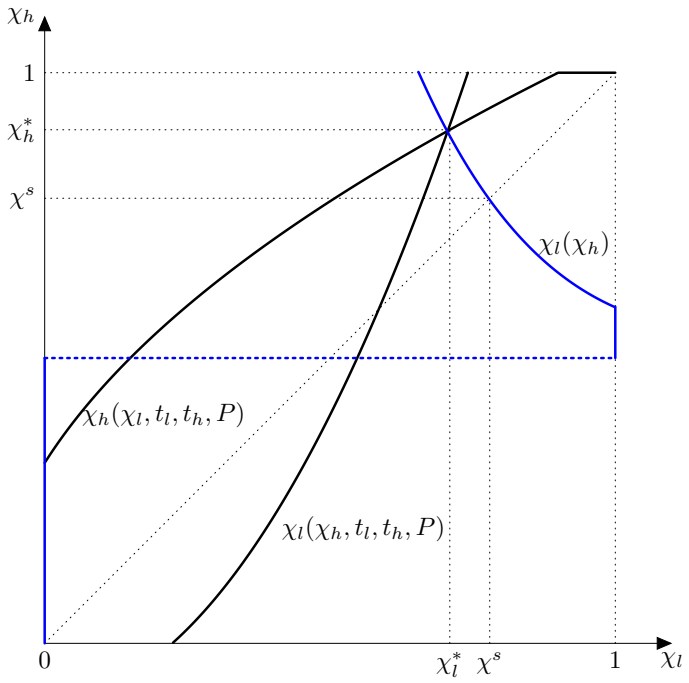

Figure 3.1 – Determination of the Electoral Equilibrium

figure). Indeed, despite the popular party capturing more rents, his valence advantage implies a higher vote share in equilibrium. This, in turn, creates a lower incentive for unpopular parties to enter the race. Consequently, less parties form when candidates are heterogeneous relative to the homogeneous and symmetric case of the first application (given that the cost of party formation stays fixed).

In sum, two interesting insights arise from this simple heterogeneous case. First, the presence of a popular party disciplines unpopular ones in that they are incentivized to capture less rents in equilibrium. Second, the high popularity party decreases the willingness of low popularity parties to enter the political market so the polity is less fragmented. Observe finally that the results will be precisely the opposite if we were to study the sequential entry of high popularity parties, given that a low popularity party is the initial entrant.

Simultaneous entry of heterogeneous parties with endogenous popularity

In this section, we assume that parties enter the race simultaneously. Furthermore, parties can choose between two formation technologies. Indeed, we posit that the resources that are invested at the formation stage allow the parties to increase their popularity. As a matter of simplification, we posit that there are only two feasible strategies. Parties can either choose to invest a high amount of resources at their formation stage. Those resources relate to media campaigns, meetings and other investments that allow candidates to gain popularity. Alternatively, parties can choose to run a low-cost campaign, in which case they enter the race with a lower popularity. Formally, we assume when a party pays a cost c_h (resp. c_l), it benefits from a popularity t_h (resp. t_l) in the election, with $c_l < c_h$ and $t_l < t_h$. In this section, we will demonstrate that the number of parties of the two types is uniquely determined, and so is the electoral equilibrium.

We denote P_l and P_h the number of parties with a low and a high reputation respectively. The equilibrium probability of elections are

$$\begin{aligned} G_h(t_l, t_h, P_l, P_h) &= \frac{t_h^K}{P_h t_h^K + P_l t_l^K Z(\chi_h, \chi_l)} \quad \text{and} \\ G_l(t_l, t_h, P_l, P_h) &= \frac{t_l^K}{P_l t_l^K + P_h t_h^K / Z(\chi_h, \chi_l)}, \end{aligned}$$

with

$$Z(\chi_h, \chi_l) = \left(\frac{1-\chi_l}{1-\chi_h}\right)^{(1-\epsilon)\theta K}. \tag{3.44}$$

Since parties maximize their expected rents, by analogy with the previous application, the optimal levels of extraction solve the following system of first-order conditions:

$$1 - \frac{1}{n}\frac{\chi_h}{(1-\chi_h)} K(1-\epsilon)\theta \left(1 - \frac{t_h^K}{P_h t_h^K + P_l t_l^K Z(\chi_h, \chi_l)}\right) = 0 \tag{3.45}$$

for the P_h high reputation parties, and

$$1 - \frac{1}{n}\frac{\chi_l}{(1-\chi_l)}K(1-\epsilon)\theta\left(1 - \frac{t_l^K}{P_l t_l^K + P_h t_h^K / Z(\chi_h, \chi_l)}\right) = 0 \tag{3.46}$$

for their P_l low reputation challengers.

Following the steps of the previous section, we can still deduce two loci $\chi_h(\chi_l, t_l, t_h, P_l, P_h)$ and $\chi_l(\chi_h, t_l, t_h, P_l, P_h)$ from the first-order conditions. Furthermore, by combining (3.45) and (3.46), we can establish a third relationship between χ_h and χ_l that proves the unicity of the intersection of the two loci (the determination of the intersection of the two loci resembles that represented in figure 3.1).

Proposition 5.

— *There exists a unique electoral equilibrium where the parties with popularity t_l (resp. t_h) extract a level of rents $\chi_l^*(t_l, t_h, P_l, P_h)$ (resp. $\chi_h^*(t_l, t_h, P_l, P_h)$), with $\chi_l^*(t_l, t_h, P_l, P_h) < \chi_h^*(t_l, t_h, P_l, P_h)$. $\chi_l^*(t_l, t_h, P_l, P_h)$ increases with t_l and decreases with t_h, while $\chi_h^*(t_l, t_h, P_l, P_h)$ decreases with t_l and increases with t_h.*

— *Both $\chi_l^*(t_l, t_h, P_l, P_h)$ and $\chi_h^*(t_l, t_h, P_l, P_h)$ decrease with P_l and P_h. Consequently, the system*

$$\begin{cases} y\chi_l^*(t_l, t_h, P_l, P_h)G_h(t_l, t_h, P_l, P_h) = c_l \\ y\chi_h^*(t_l, t_h, P_l, P_h)G_l(t_l, t_h, P_l, P_h) = c_h \end{cases} \tag{3.47}$$

admits a unique solution $(P_l^*(t_l, t_h, c_l, c_h), P_h^*(t_l, t_h, c_l, c_h))$. The aggregate number of parties is such that $P_h^S < P_l^* + P_h^* < P_l^S$, with P_l^S (resp. P_h^S) the number of parties in a symmetric equilibrium where there are only low (resp. high) types running for the election.

— *Allowing political parties to make high campaign investments decreases party fragmentation and increases rent extraction from popular parties. Furthermore, P_l^* increases with t_l and c_h while it decreases with t_h and c_l. Alternatively, P_h^* decreases with t_l and c_h while it increases with t_h and c_l.*

Démonstration. The proof is available in Appendix 3.6.6. □

The intuition for the first point of the proposition is precisely the same as in the previous application. By combining the two first-order conditions, we can establish a third relationship between χ_l and χ_h, which particularly simple form proves the unicity. Again, the complementarity between χ_l^* and χ_h^* along the equilibrium path is always dominated and an increase in t_h will lead to higher rents captured by the popular party and lower rents captured by its challengers. Similarly, when either P_h or P_l increases, then the competition becomes fiercer and parties decrease their level of extraction. We also establish in Appendix 3.6.6 that the vote share of any party is decreasing in both P_l and P_h, meaning that a new entrant of any type will decrease the vote share of his challengers.

In order to give some intuition for the unicity of the electoral equilibrium, we have represented the first line of (3.47) in figure 3.2. We see that since $y\chi_l^*(t_l, t_h, P_l, P_h)vs_l(t_l, t_h, P_l, P_h)$ is a decreasing function of P_l, it intersects only once the line c_l. Furthermore, the intersection defines a locus $P_l(P_h, t_l, t_h)$ that is decreasing with P_h, since $y\chi_l^*(t_l, t_h, P_l, P_h)vs_l(t_l, t_h, P_l, P_h)$ decreases with P_h. Consequently, (3.47) characterizes two loci $P_l(P_h, t_l, t_h, c_l)$ and $P_h(P_l, t_l, t_h, c_h)$ that we have represented in figure 3.3. The unicity of the electoral equilibrium follows from monotonicity of $P_l(P_h, t_l, t_h)$ (resp. $P_h(P_l, t_l, t_h)$) with respect to P_h (resp. P_l).

Relative to the comparative statics, when t_l decreases for instance, then the high popularity parties are incentivized to capture more rents because they face weaker challengers. Low popularity candidates alternatively decrease their rents in order to provide high quality platforms. Furthermore, high popularity candidates get higher vote shares relative to a symmetric equilibrium with high valence candidates. Indeed, assume that parties initially have a high valence and some of them (a fraction $P_l/(P_l + P_h)$) see their valence decrease from t_h to t_l. High popularity parties increase their level of extraction along the equilibrium path in such as way to also increase their vote share in equilibrium, given the reaction of low popularity candidates. In other words, the increase in χ_h does not entirely compensate their valence advantage. Consequently, the value function of high popularity candidates necessarily increase when t_l decreases and so more high popularity candidates should enter the race. By contrast, a lower value of t_l will deter low popularity candidates from entering the race. The

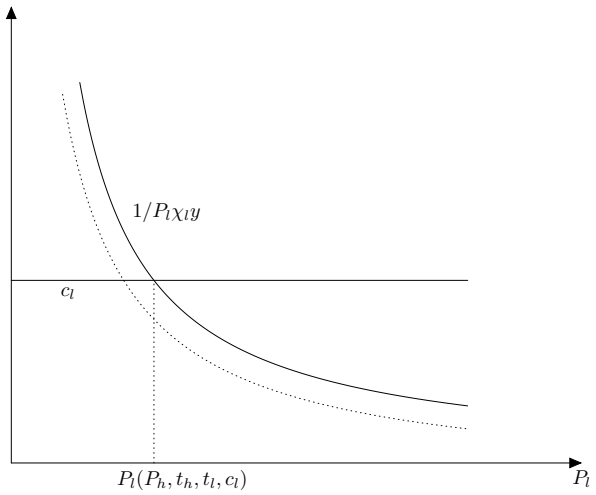

Figure 3.2 – Determination of the locus $P_l(P_h, t_l, t_h, c_l)$ from the system (49) and effect of an increase in P_h on $P_l(P_h, t_l, t_h, c_l)$ (dotted curve).

intuitions for the rest of the comparative statics are similar.

It is finally interesting to observe that in equilibrium, we should expect the aggregate number of parties to be bounded between the equilibrium number of parties in a symmetric equilibrium with only high popularity candidates (lower bound) and low popularity candidates (higher bound). Of particular interest, this result implies that the presence of high popularity candidates deters low popularity candidates from entering the race relative to the case where high investments during electoral campaigns are not allowed. Furthermore, allowing parties to make high campaign investments aiming at increasing their popularity will lead to higher rent extraction in equilibrium, for the simple reason that a higher valence decreases candidates' incentive to transfer resources to the citizenry.

In sum, electoral competition limits but is not sufficient to completely suppress rent extraction from the candidates able to benefit from significantly higher fundings than their peers in electoral campaigns and high campaign investments decrease party fragmentation although they do not completely deter the entry of small candidates.

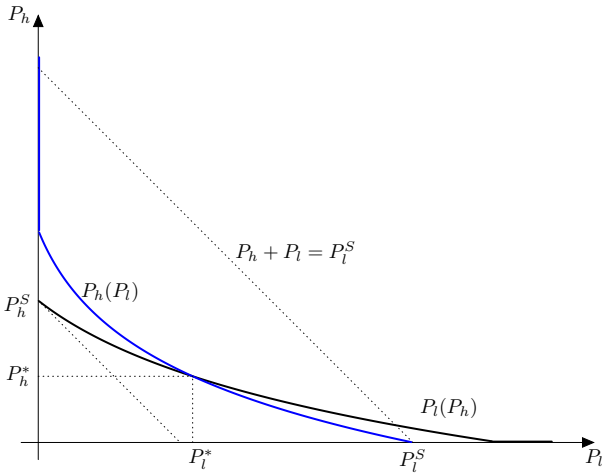

Figure 3.3 – Determination of the electoral equilibrium with endogenous entry of heterogeneous candidates.

3.4.3 Application 3: Endogenous entry and coalition formation in the legislature in proportional systems

This section studies the formation of coalitions in the legislature under proportional electoral rules. Arguably, the possibility to form coalitions in the legislature should affect parties' incentive to run as well as their political platforms. As a simple illustration, we should expect the incentive of small parties to run in proportional elections to be higher despite their low expected seat shares in the legislature when they anticipate that they might be able to participate to governing coalitions. This application will show that this intuition is not entirely true. Besides, to our knowledge, the "first stage" of the coalition formation game in the legislature that accounts for the formation of parties has not been apprehended yet in the literature. This application fills this gap.

We assume that the parties get a fraction of the seats in the legislature that equalizes their vote share in a single election. We do not introduce thresholds for simplicity. One party is chosen in the legislature to form a government. This party is labeled the formateur hereafter. We assume that the probability that a given party is chosen to form a governing coalition is equal to its seat share in the legislature. This assumption has been made by Baron and Diermeier (2001) for instance, and finds empirical support in the analysis of Diermeier and

Merlo (2004). If the formateur does not manage to form a government, then each legislator gets some exogenous rents r and the game ends. We posit that those rents are independent from the seat share of the parties in the legislature. Once chosen, the formateur makes a take-it-or-leave-it offer to some other parties in the legislature so as to form a minimum winning coalition. A legislator will therefore accept to be in the governing coalition if the formateur makes an offer that is above and arbitrarily close to r.

We focus on the case of special interest politics with heterogeneous voters and homogeneous candidates (see the first application). The set of individual attributes $s \in S$ characterize a set of interest groups. The utility of the agents in the interest group $s \in S$ only depends on their consumption that we denote c_s. Party $i \in P$ transfers $\tau_{s,i}$ units of income to any agent with attributes $s \in S$ and extracts a fraction $\chi_i \in [0,1]$ of the tax base as rents. The strategy of a party will consist in setting a policy vector $q_i = \{\chi_i, \{\tau_{s,i}\}_{s\in S}\}$ and in building a minimum winning coalition C_i in case it is chosen to be the formateur. C_i is a subset of the set of parties in the legislature that we still denote P. The objective of party i can be written in the form

$$\max_{C_i,q_i} W(C_i,q_i) = vs_i(q_i,C_i,q^*_{-i},C^*_{-i})[\chi_i y - rP \sum_{j\neq i, j\in C_i} vs_j(q^*_j,C^*_j,q_{-j},C_{-j})] + \sum_{j\neq i} vs_j(q^*_j,C^*_j,q_{-j},C_{-j})\Pr(i\in C^*_j) rP vs_i(q_i,C_i,q^*_{-i},C^*_{-i}), \quad (3.48)$$

under the constraints

$$\begin{cases} -\sum_{s\in S} x_s \tau_{s,i} \geq \chi_i \overline{y} + rP \sum_{j\neq i, j\in C_i} vs_j(q^*_j,C^*_j,q_{-j},C_{-j}) \\ \sum_{j\neq i, j\in C_i} vs_j(q^*_j,C^*_j,q_{-j},C_{-j}) + vs_i(q_i,C_i,q^*_{-i},C^*_{-i}) \geq 1/2 \\ \max_{C_i,q_i} W(C_i,q_i) \geq \max_{C_i,q_i} rP vs_i(q_i,C_i,q^*_{-i},C^*_{-i})\{1+\sum_{j\neq i} vs_j(q^*_j,C^*_j,q_{-j},C_{-j})\Pr(i\in C^*_j)\}, \end{cases} \quad (3.49)$$

where q^*_j and C^*_j are respectively the optimal policy and the minimum winning coalition of party j. The vector q_{-j} is such that $q_{-j} = \{q^*_1,\ldots,q_i,\ldots,q^*_{j-1},q^*_{j+1},\ldots,q^*_P\}$. Similarly, C_{-j} denotes the vector of optimal coalitions of j's challengers given that i does not have an optimal coalition.

The first constraint is a standard budget constraint. The formateur extracts resources from the citizenry, takes some rents and transfers r units of revenues to a set of legislators that belongs to the parties in the coalition C_i. The second constraint says that a minimum winning coalition C_i is such that the seat share of C_i is at least equal to one half. Indeed, assume that the policies are enacted by majority voting in the legislature. It is strictly suboptimal for the legislators that are not part of the governing coalition to vote for the platform of the formateur because it incentivizes him to buy their support. Furthermore, observe that the formateur will not make offers below r in equilibrium, precisely because any party that belongs to the coalition is median in that if it leaves, then the coalition loses the majority in the legislature. Finally, there is still an incentive compatibility constraint, since building a minimum winning coalition should provide higher rents. Indeed, it could be that the formateur prefers not to build a coalition, in which case he gets an exogenous level of rents r. This incentive compatibility constraint is described in the third line of (3.49). Observe nevertheless that if r is sufficiently low relative to the perks of passing a tax policy in the legislature, then this outcome is unlikely. We will posit hereafter that the third constraint is always respected, so that formateurs prefer to form coalitions.

Relative to the objective (3.48), if party i is chosen to be the formateur with probability vs_i - and given that it builds a minimum winning coalition, it will be able to set the tax policy and to fix the level of rents to χ_i. Furthermore, an amount r will be transferred to the $rP\sum_{j\neq i, j\in C_i} vs_j$ legislators. Relative to the second line of (3.48), with probability vs_j, party j is chosen to be the formateur, in which case the Pvs_i legislators of party i receive a transfer r if they are chosen to be part of j's governing coalition. We denote $\Pr(i \in C_j^*)$ the probability that i is chosen to be part of C_j^*, for $j \neq i, j \in P$

For simplicity, we will assume that the parties have the same popularity, i.e. $t_i = t_j$ for any $i, j \in P$. Under those conditions, parties are homogeneous. We moreover focus on the determination of the symmetric electoral equilibrium, if it exists, since the global concavity of any party's objective can not be obtained in a simple way.

In a symmetric electoral equilibrium, parties have the same vote share and thus any

party $j \in P$ has an equal chance of participating to a coalition initiated by any party $i \in P$. As before, the number of parties will be the largest integer below a real number P that is endogenously determined. We denote $E(P)$ the largest integer below P and assume that there are at least two parties represented in the legislature. In those conditions, it is direct that

$$\Pr(i \in C_j^*) = \binom{E(P)-2}{E(E(P)/2)-2} \Big/ \binom{E(P)-1}{E(E(P)/2)-1}. \tag{3.50}$$

The denominator gives the total number of subgroups of size $E(E(P)/2)-1$ in a set of size $E(P)-1$. Thus, the denominator of (3.50) gives the number of possible coalition of size $E(E(P)/2)$ that includes party j. It represents the number of minimum winning coalitions that include a given party j. By analogy, the numerator is the set of winning coalitions that include both j and i. Indeed, if i belongs to the coalition formed by party j, then party j still has to choose $E(E(P)/2)-2$ other parties for the coalition among the set of remaining $E(P)-2$ parties. $\Pr(i \in C_j^*)$ simplifies to

$$\Pr(i \in C_j^*) = \frac{E(P)/2-1}{E(P)-1}. \tag{3.51}$$

Combining (3.48) and (3.49), and using the symmetry assumption, we can rewrite the objective of party i as

$$\begin{aligned}\max_{C_i,q_i} W(C_i,q_i) = {} & vs_i(q_i,C_i,q_{-i}^*,C_{-i}^*)[\chi_i y - rP(1/2 - vs_i(q_i,C_i,q_{-i}^*,C_{-i}^*)]+ \\ & (1-vs_i(q_i,C_i,q_{-i}^*,C_{-i}^*))\frac{E(P)/2-1}{E(P)-1} rP vs_i(q_i,C_i,q_{-i}^*,C_{-i}^*).\end{aligned} \tag{3.52}$$

We assume that P is an even integer for simplicity in the sequel. In order to gain intuitions on the mechanisms at play, we follow the steps of the first application and write the first-order condition with respect to the transfers $\tau_{s,i}$:

$$-\frac{1}{P}n_s + (1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s}\{\chi_i y + \frac{1}{2}r\frac{P}{P-1}\}\frac{P-1}{P^2} = 0. \tag{3.53}$$

Increasing the transfers to the interest group s still decreases the level of extraction (first term in the LHS of (82)). Relative to the first application however, the marginal benefits from increasing the transfers to group s changes. Indeed, when party i increases marginally its vote share, (i) it increases the likelihood of being the formateur (ii) reduces marginally the cost of building a winning coalition and (iii) increases the transfers that can be gained from participating to a governing coalition without being the formateur. This is why the second term in the LHS of (3.53) is higher than in the first application, since when there is no coalition formation, only effect (i) is at play. As in the first application, it follows immediately from the first-order condition that

$$\theta_s/c_s = \theta_k/c_k \tag{3.54}$$

for any pair of interest group $s, k \in S$. Using the budget constraint then, we deduce that

$$c_s = \frac{\theta_s}{n\overline{\theta}}(y(1-\chi_i) - r(\frac{P}{2} - 1)). \tag{3.55}$$

Not surprisingly, since parties have to invest resources in order to build a winning coalition, the consumption level c_s is decreasing in the number of parties in the legislature and in the exogenous rents r that have to be redistributed to the parties in governing coalitions. Observe then from (3.53) that we should expect the marginal benefit of transferring resources to any interest group s to be larger than in the case of the first application because consumption levels are anticipated to be lower. Following the steps of the first application, we can deduce from this point the level of rent extraction χ_i and the number of parties P_c^* in equilibrium. The computations are detailed in Appendix 3.6.7. [11]

Proposition 6. *In equilibrium, the number of parties is the largest integer below P_c^*, with*

$$P_c^* = \frac{1}{1 + (1-\epsilon)\overline{\theta} + r/2c}[(1-\epsilon)\overline{\theta} + \frac{y + r(1 - \overline{\theta}/2)}{c}]. \tag{3.56}$$

11. It is interesting to note that in a symmetric equilibrium, the expected cost of forming a coalition is equal to the expected benefit of participating to a coalition without being the formateur. Indeed, in a symmetric equilibrium. One can verify that (80) simplifies to $W(C_i, q_i) = \frac{1}{P}\chi_i y$.

The comparative statics of the first application are robust to the formation of coalitions. Furthermore, the number of parties decreases with the rents of legislators r. The formation of coalitions in the legislature reduces party fragmentation and there are strictly more parties under plurality than proportional electoral rules.

Démonstration. The proof is available in Appendix 3.6.7. □

Our analysis of stochastic and sincere voting contradicts Duverger's prediction that proportional systems should favor higher party fragmentation. At the center of our prediction lies the interaction between public policy, coalition formation and party formation. In particular, introducing coalitions creates stronger incentives for parties to increase their vote share (and thus to decrease rent extraction). Furthermore, the cost of coalition formation is partially born by the citizenry as it lowers the feasible transfers (and in turn strengthen even more parties' incentive to reduce their extraction level). To conclude with an illustration, take the case where $r/c = 3$, $y/c = 5$ and $\overline{\theta} = \epsilon = 0.5$. In that case, we should expect only 2 parties to form under the proportional system, while in a plurality system, there should be as much as 6 parties from Proposition 2.

3.4.4 Application 4: Redistributive policies under alternative voting rules

The Duverger's hypothesis, as formulated by Riker (1982), states that runoff elections should produce more candidates than plurality elections. Furthermore, the Duverger's law states that simple majority single ballot plurality favors the two party system whereas simple majority with a second ballot or proportional representation favors multipartysm (Riker (1982)). In this application and the next, we intend to study whether our theory of stochastic and sincere voting is consistent with Duverger's hypothesis.

This application compares the endogenous entry of political parties in runoff and plurality elections. The case of plurality elections has been developed in the canonical model of Section 3.2 and applied to redistributive politics with homogeneous candidates in Section 3.4.1. The

runoff system by contrast proceeds in two rounds. In the first round, the two candidates with the highest vote shares are selected and are allowed to run for the second round. This is the system used in the French presidential election for example. We will first develop a general framework for the study of runoff elections and then apply it to a simple case of redistributive politics with homogeneous candidates and heterogeneous voters. As a matter of simplification, we assume that in runoff elections, parties promise platforms before the first round and can not back pedal on their promises between the two rounds and citizens have to show up to the ballot twice.

Let $ij(q_i, q_{-i})$ be the probability that the citizenry ranks i and j first given the list of possible candidates P. Using the results of the first application, we know that the fraction of voters that cast a ballot for candidate i is given by (27):

$$vs_i(q_i, q_{-i}) = vs_i^t(q_i, q_{-i})\eta(i, P),$$

with

$$vs_i^t(q_i, q_{-i}) = \sum_{s\in S} \frac{n_s}{n} \frac{t_i u(c_s(q_i))^{\theta_s}}{t_i u(c_s(q_i))^{\theta_s} + \sum_{j\in P\setminus i} t_j u(c_s(q_j))^{\theta_s}}.$$

We deduce that $ij(q_i, q_{-i})$ can be expressed in the form

$$ij(q_i, q_{-i}) = \Pr[\min(vs_i^t(q_i, q_{-i})\eta(i, P), vs_j^t(q_j, q_{-j})\eta(j, P)) > vs_k^t(q_k, q_{-k})\eta(k, P) \text{ for any } k \in P \setminus \{i, j\}].$$

Using the same expressions as in the first application for the probabilities $\eta(i, P)$, $i \in P$, we show the following result in the Appendix.

Lemma 2. *The probability that i and j are ranked in the two first positions in the first round of a runoff election can be expressed as*

$$ij(q_i, q_{-i}) = \frac{{vs_i^t}^K}{\sum_{k\neq j} {vs_k^t}^K} + \frac{{vs_j^t}^K}{\sum_{k\neq i} {vs_k^t}^K} - \frac{{vs_i^t}^K + {vs_j^t}^K}{\sum_k {vs_k^t}^K}. \tag{3.57}$$

Démonstration. The proof is available in Appendix 3.6.8. □

Lemma 2 follows from a standard result in probability theory on the distribution of the minimum of a list of random variables, and from the property that the product of two Fréchet cdf is also a Fréchet cdf. The probability $P_i(q_i, q_{-i})$ that i wins the runoff election can finally be expressed as

$$P_i(q_i, q_{-i}) = \sum_{j \neq i} ij(q_i, q_{-i}) [\frac{\tilde{vs}_i^t(q_i, q_j)^K}{\tilde{vs}_i^t(q_i, q_j)^K + \tilde{vs}_j^t(q_j, q_i)^K}], \tag{3.58}$$

with $\tilde{vs}_i^t(q_i, q_j)$ the "theoretical" vote share of party i in a pairwise election against j. The bracketed term in (3.58) gives the probability that i wins a pairwise election against candidate j, see Section 2.1. One can finally observe that in a symmetric equilibrium, the probability that i is elected simplifies to $1/P$.

We will now apply the model to the case of special interest politics with heterogeneous voters and homogeneous entrants and focus on the symmetric electoral equilibrium. Assume that the utility function of the agents in group s is given by (3.20). The platform of any party i consists in a vector of transfers $\{\tau_{s,i}\}_{s \in S}$ and a level of extraction χ_i. In those settings, we were able to determine a closed form result for the number of parties in plurality elections in the first application,

$$P_p^* = \frac{1}{1 + \bar{\theta} K (1 - \epsilon)} [\bar{\theta} K (1 - \epsilon) + \frac{y}{c}]. \tag{3.59}$$

Consider now the case of runoff elections. In an electoral equilibrium, i's platform solves

$$q_i^* = \arg\max_{q^i} P_i(q_i, q_{-i}^*) \chi_i y, \tag{3.60}$$

with

$$\chi y \leq - \sum_{s \in S} n_s \tau_{s,i}. \tag{3.61}$$

We show in Appendix 3.6.9 that in a symmetric equilibrium, the optimal level of rents

extracted by political parties in runoff elections is determined by the following condition:

$$-\frac{1}{P}+(1-\epsilon)\overline{\theta}K\frac{\chi_m}{1-\chi_m}[\frac{1}{2}\frac{(P-2)(2P-1)}{P^2(P-1)}+\frac{1}{2}\frac{1}{P}]=0. \tag{3.62}$$

As a matter of comparison, recall that in the case of plurality elections, the condition that determines the level of extraction writes

$$-\frac{1}{P}+(1-\epsilon)\overline{\theta}K\frac{\chi_p}{1-\chi_p}[\frac{P-1}{P^2}]=0. \tag{3.63}$$

Thus, when increasing marginally the transfers to the interest groups, party i decreases its rents by a corresponding amount in the two systems. The marginal benefits from doing so are alternatively different. In plurality elections, from (3.62), by increasing marginally the transfers, party i grabs a fraction $1/P^2$ of the vote share of each of its $P-1$ challengers, abstracting from the effects of ϵ, $\overline{\theta}$ and of the level of extraction χ_p. The first and the second terms in the bracket in (3.62) reflect respectively the effect of increasing the transfers on party i's vote share in the first round and in the second round. Observe first that the probability of being in the pair of candidates that passes the first round is proportional to $P-2$, since a pair faces $P-2$ challengers. It is also proportional to one half, the probability of being elected in the second round in a symmetric equilibrium. The term $(2P-1)/(P^2(P-1))$ reflects the marginal probability of being in the pair of candidates selected for the second round. In order to interpret the second term in the bracket, notice that it is also equal to the product of $1/4$ and $(P-1)(2/(P-1)-2/P)$. Indeed, $1/4$ is the marginal vote share in a pairwise election and $(P-1)(2/(P-1)-2/P)$ is the probability of being in a pair of candidates selected for the second round of the election in a symmetric equilibrium.

Taken separately, the first or the second rounds induce lower marginal benefits from transferring resources to the citizenry relative to plurality elections. Indeed, it can easily be

shown that when $P \geq 2$, then

$$\begin{cases} (P-2)(2P-1)/(2P^2(P-1)) \leq (P-1)/P^2 \text{ and} \\ 1/(2P) \leq (P-1)/P^2, \end{cases} \tag{3.64}$$

meaning that each round in a runoff election incentivizes *less* the parties to transfer resources to the citizenry. Indeed, when comparing the first round of a runoff election and the plurality election, it is clear that in the former case, parties face less competition (and similarly for the second round as long as $P \geq 2$). Interestingly however, taken together, the first and the second rounds of a runoff election creates higher marginal benefits from transferring resources to the citizenry, since

$$\frac{1}{2}\frac{(P-2)(2P-1)}{P^2(P-1)} + \frac{1}{2}\frac{1}{P} \geq \frac{P-1}{P^2} \tag{3.65}$$

when $P \geq 2$. In other words, the repetition of electoral competition in two consecutive rounds dominates the lower electoral competition that parties face in each round taken separately relative to plurality elections. In sum, this also implies that parties should capture *less* rents under the runoff system than under the plurality system. The direct consequence of this is that *less* parties should form in the runoff system, assuming that the cost of party formation is independent of the electoral rule. The following result summarizes the previous discussion.

Proposition 7.

— *The equilibrium number of parties in a runoff electoral system P_m^* is uniquely determined by the equation $1/P\chi_m y = c$, with χ_m a decreasing function for $P \geq 1$ such that*

$$\chi_m = \frac{1}{1 + \overline{\theta}K(1-\epsilon)/2\{1 + (P-2)(2P-1)/(P(P-1))\}}. \tag{3.66}$$

— *With stochastic and sincere voting, the Duverger's hypothesis never holds and $P_m^* > P_p^*$ necessarily.*

Démonstration. The proof is available in Appendix 3.6.9. □

In a model with sincere and stochastic voting, the difference between runoff and plurality

elections pins down to the effect of the two electoral systems on the marginal benefits of parties from capturing rents (see (3.62) and (3.63)). Taken separately, the two rounds of a runoff elections create a stronger incentive for parties to capture rents, because the competition is lowered relative to plurality elections. But taken together, the two rounds creates a weaker incentive for parties to capture rents and this explains why we should expect less parties in equilibrium to form under runoff systems.

We illustrate the determination of the number of parties in the two electoral systems in figure 3.4 in the case where $y = 1.5$, $c = 0.1$ and $(1 - \epsilon)\overline{\theta} = 0.95$. In this example, we should expect respectively 8 and 6 parties in plurality and runoff systems. To conclude, this section shows that stochastic and sincere voting does not generate Duverger's outcome but the precise opposite.

3.4.5 Application 6: Peaceful franchise extension in Democracy

A key prediction of theories where elections determine the allocation of resources is that increased political participation of an interest group leads to more redistribution toward them (e.g. Meltzer and Richard (1981), Coughlin and Nitzan (1981), Lindbeck and Weibull (1987), Coughlin (1992), Dixit and Londregan (1996)). A direct correlate of this prediction is that the enfranchisement of the poor should lead to more redistribution (Acemoglu and Robinson (2000)). [12] The existing evidence is mixed however on the effect of democratization on government spendings. Aidt and Jensen (2009a,b, 2013), Acemoglu, Naidu, Restrepo, and Robinson (2015) and Fujiwara (2015) find positive effects for example, while Mulligan, Gil, and i Martin (2004) argue that there are no differences in policymaking in autocracy and democracy.

In this extension, we extend the model of the first application in order to account for a peaceful process of franchise extension given that the fragmentation of the polity is endogenous. In doing so, we will show why it is reasonable to expect public spendings to increase when the franchise is extended, *independently* from the preferences of the citizenry. We will explain

12. See as well Lizzeri and Persico (2004) for an alternative theory on the enfranchisement of the poor.

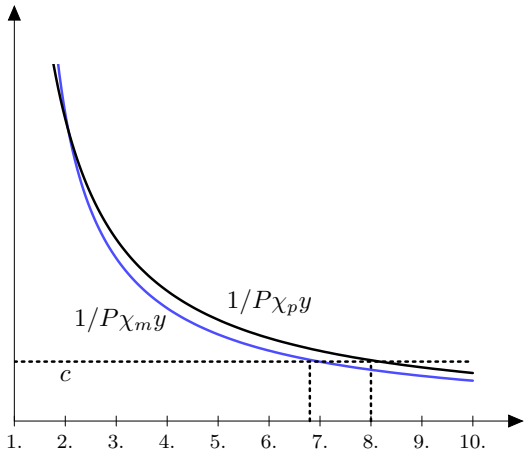

Figure 3.4 – Determination of the number of parties in plurality elections (black curve) and runoff elections (blue curve).

as well under which conditions should the extension of the franchise be supported by the citizenry.

We assume that the set of interest groups S is divided in two subsets S_f and S_n. If an individual belongs to $s \in S_f$, then he is allowed to vote, while if $s \in S_n$, he can not cast a vote. As in the first application, we restrict the study to special interest politics with heterogeneous voters and homogeneous entrants. The transfers satisfy the following budget constraint,

$$y_n - \sum_{s \in S_f} n_s \tau_{s,i} \geq (y_f + y_n)\chi_i, \tag{3.67}$$

with χ_i the fraction of the revenues captured as rents by party i, $y_f = \sum_{s \in S_f} n_s y_s$ and $y_n = \sum_{s \in S_n} n_s y_s$. Indeed, the revenues of the disenfranchised are entirely extracted by the parties and a share χ_i is extracted as rents.

Following the steps of the first application, we establish in Appendix 3.6.10 that

$$P^* = \frac{1}{1 + K\overline{\theta}_f(1-\epsilon)}[K\overline{\theta}_f(1-\epsilon) + \frac{y_f + y_n(1-\alpha_0)}{c(1-\alpha_0)}], \tag{3.68}$$

and

$$c_s^* = \frac{K\theta_s(1-\epsilon)}{n_f(1 + K\overline{\theta}_f(1-\epsilon))}(y_f + (1-\alpha_0)y_n - c(1-\alpha_0)) \tag{3.69}$$

for any $s \in S_f$, with $\overline{\theta}_f = \sum_{s \in S_f} n_s \theta_s$.

Note that the formula corresponds to precisely to those established in the first application (3.36) and (2.48) up to two important alterations. First, only the average responsiveness of the enfranchised agents matters. This is straightforward, since the disenfranchised are by definition not authorized to vote - and pose no threat of revolution - their political characteristics do not matter in the setting of public policies. Second, both the equilibrium number of parties P^* and the vector of consumption depend on the maximum aggregate consumption of the enfranchised $y_f + (1-\alpha_0)y_n$. Indeed, if all the citizens were allowed to vote, then the maximum aggregate consumption of the enfranchised will simply be y_f. When there are no transfers (and thus no leaks), there is no rent extraction so the aggregate consumption is equal to y_f. By contrast when a fraction of the citizenry is not allowed to vote and their revenues y_n are extracted, then the maximum feasible consumption of the enfranchised equalizes the sum of y_f and the share of the revenues of the disenfranchised that is not lost in the process of extraction $(1 - \alpha_0)y_n$.

An extension of the franchise corresponds formally to a switch of some group $s \in S$ from S_n to S_f. That is, n_f increases by n_p if group $p \in S$ is enfranchised, while n_n decreases by n_p.

Observe first that when the disenfranchised are passive, i.e. they pose no threat of revolution and when the parties and the enfranchised interests have no vested interests - as it is assumed here - then political parties should be indifferent to the issue of franchise extension. Indeed, even if the enfranchisement of some interest groups allows to capture more rents in the short term, this will necessarily induce more entry on the political arena until parties' expected rents converge back to the cost of party formation c.

Proposition 8.

— *The extension of the franchise creates an income effect that increases the scope of public spendings. If an interest group with a political responsiveness θ_c is allowed to vote, then the share of public spendings transferred to this group is strictly higher (resp. lower) than in the disenfranchised case if θ_c is higher (resp. lower) than some threshold $\hat{\theta}_f$.*

— *The citizenry has a relatively higher incentive to extend the franchise to interest groups with a low political responsiveness. Franchise extension is relatively more likely when the fiscal capacity is low and when income disparities between the enfranchised and the disfranchised agents are low.*

Démonstration. The proof is available in Appendix 3.6.10. □

Before detailing the intuition of this result, observe from (3.69) that the consumption of any enfranchised group s depends multiplicatively on θ_s and on a function of the *average* characteristics of the citizenry. In this model then, all the enfranchised interests have the same incentive to extend the franchise, since what matters to them is how the *average* characteristics of the citizenry are affected by the extension of the franchise. Our theory is then different of that of Lizzeri and Persico (2004) in that respect, since enfranchisement creates here no conflict between the enfranchised interests. Furthermore, our result implies that the absence of vested interests between political parties and the enfranchised groups should necessarily lead to an extension of the franchise to non-politically responsive groups, especially when the fiscal capacity is low.

The results of proposition 8 obtain because (i) the extension of the franchise necessarily creates an income effect that is stronger when the fiscal capacity is low and (ii) allowing non-responsive agents to vote decreases the marginal cost of redistributing resources to pre-existing interest groups. The claim that an extension of the franchise induces an income effect has not been made before in the literature to our knowledge, yet it follows from a simple inspection of the evolution of the government's budget constraint as we explain below.

When the franchise is extended, the maximum aggregate consumption $y_f + (1 - \alpha_0)y_n$ necessarily increases for $\alpha_0 > 0$. That is to say, the enfranchisement of new interest groups induces a positive income effect since it reduces the extent of inefficiencies in the setting of public policies. Take the simple example of a franchise extended to a single citizen c. If c is not allowed to vote, then his income y_c is fully extracted and a fraction $(1 - \alpha_0)$ is redistributed to the citizenry in the best case scenario (i.e. when the parties capture no rents). Thus, interest

groups expect to consume in equilibrium a share of the total revenues $y_f + (1 - \alpha_0)y_c$ that is proportional to their relative political responsiveness. If c is allowed to vote, then the interest groups expect to consume a share of $y_f + y_c > y_f + (1 - \alpha_0)y_c$. This is because parties have the possibility of not extracting resources from agent c. The feasibility of such a platform makes policies inherently more efficient *for any* citizen.

The preceding income effect explains why it is reasonable to expect public spendings to increase when the franchise is extended. Consistently with this result, series of evidence suggest that public spendings increase when the franchise is extended, e.g. Aidt and Jensen (2009a,b, 2013), Acemoglu, Naidu, Restrepo, and Robinson (2015) and Fujiwara (2015). It is nevertheless not guaranteed that those higher spendings will be directed toward the new enfranchised groups. On the contrary, if the newly enfranchised groups have a low political responsiveness, then we should expect the benefits of a peaceful enfranchisement to be disproportionately directed toward the pre-existing interest groups while the newly enfranchised still benefit at the margin as a consequence of the income effect induced by the extension of the franchise.

Finally, the income effect is stronger when the fiscal capacity is low. Indeed, if some group $p \in S$ is allowed to vote, then the aggregate revenues increase by $\alpha_0 n_p y_p$. Since an extension of the franchise is more likely when the fiscal capacity is low, we should expect fiscal consolidation to be preceded by franchise extension and not the contrary.[13] Note finally that wealthier enfranchised groups have less incentives to further the enfranchisement, since the standard dilution effect implied by the extension of the franchise is stronger when y_f is high. Thus, we should expect income inequality to limit the scope of democratization. This accords with the study of Engerman and Sokoloff (2005) for example, that find support that greater equality generally led in broadening the franchise within the United States. More broadly, Engerman and Sokoloff (2000) suggest that the late extension of the franchise in Latin America reflects initial differences in economic structures and wealth disparities relative to the United States and Canada. Observe however that our prediction that income inequalities decrease the willingness of the elite to extend the franchise depends crucially on the assumption that the

13. The incentive of political parties in consolidating their fiscal capacity is beyond the scope of this theory. On this subject, see the review of Besley and Persson (2013) and the references therein.

disenfranchised interests pose no threat to the political institutions, which arguably could be higher when income disparities are large, as emphasized by Acemoglu and Robinson (2000).

3.5 Conclusion

We have provided in this paper a new stochastic voting model for multi-candidate elections. Indeed, we have demonstrated that Fréchet (or extreme type II) distributions significantly ease the issue of computing candidates' objective functions in plurality, run-off and proportional elections when the number of candidates is arbitrary. Relying on an axiomatic approach to probabilistic voting theories inspired of McFadden (1974), we were able to micro-found the use of Fréchet distributions in stochastic voting models.

Our central objective in this paper was to show that our theory is flexible and provides a unifying framework to study various topics of the political economy literature that have traditionally grown independently. Five applications of the canonical model are developed on the topics of special interest politics, public good provision versus redistribution, Duverger's law and hypothesis, coalition formation in the legislature and franchise extension in democracy. Those applications yield various refutable predictions that are provided in the propositions. Among the key contributions of the paper, we establish that both proportional and run-off systems should lead to higher party fragmentation than a plurality system. Furthermore, we were able to show that allowing high campaign investments leads to less fragmented polities and more rent extraction from well funded political parties.

Important issues have been left aside, which could nevertheless be approached with the stochastic model presented in this paper. For instance, we have not apprehended the issue of the formation of opposition movements in autocracies. Arguably, a theory on the formation of an endogenously fragmented opposition could help researchers understand strategies of entrenched elites willing to avoid a democratic transition. [14] Furthermore, it could be particularly interesting to extend this theory in order to account for more complex

14. This issue has been studied by Acemoglu, Verdier, and Robinson (2004) and Padró i Miquel (2007) for example.

technologies of party formation and study the process of selection of candidates within parties (e.g. Caillaud and Tirole (2002)). Finally, the issue of the policy motivations or ideologies (e.g. Snyder and Ting (2002)) of parties and candidates has been left aside in this paper, since we have focused on the effect of political competition on the fragmentation of the polity. Future works may account for both policy and office motivations. Such an extension could ultimately help disentangle the effect of preferences from that of pure competition on the motives of political entry under various electoral rules.

3.6 Appendix

3.6.1 Proof of Theorem 1

Démonstration. Existence: Consider the function

$$h_{si}(y) = \frac{t_i y}{t_i y + K_i} \tag{3.70}$$

defined on a convex space E_i that contains $V(s, Q_i)^{\theta_s}$ (such a space exists since Q_i is convex and $V(s, .)$ is continuous) is straightforwardly concave.

As the vote share of any candidate i can be expressed as

$$vs_i(q_i, q_{-i}) = \sum_{s \in S} h_{si}(V(s, q_i)^{\theta_s}), \tag{3.71}$$

$vs_i(., q^{-i})$ is concave on $Q_i \subseteq Q$ whenever $q_i \to V(s, q_i)^{\theta_s}$ is concave for any s. Finally, since $vs_i(., .)$ is continuous on $\prod_{j \in P} Q_j$, we can apply a standard result of equilibrium existence (e.g. Fudenberg and Tirole (1991, p. 34)) that we state below.

Theorem. *(Adapted from Fudenberg and Tirole (1991, p. 34) to fit the concept of electoral equilibrium given in Definition 1). Consider a strategic-form game whose strategic spaces Q_i are non-empty compact convex subsets of an Euclidian space. If the vote share functions vs_i are continuous in $\prod_{j \in P} Q_j$ and quasi-concave in Q_i there exists a pure-strategy voting*

equilibrium.

Unicity: Whenever there exists $s \in S$ such that $V(s,.)^{\theta_s}$ is strictly concave over the set of feasible policies Q, then it follows that the vote share of any candidate i in the subset of voters with attributes $s \in S$ is strictly concave on q_i. Thus, the vote share of any candidate i is strictly concave over Q_i, which implies unicity. □

3.6.2 Proof of Theorem 2

This theorem is again inspired of the seminal approach of Daniel McFadden, and more specifically the first two Lemma of McFadden (1974) although it extends a bit his result to account for some heterogeneity in the distributions of the noise parameters.

Assume first that the agents use a random utility voting model with a noise distributed according to Fréchet distributions $F_{i,s}(\epsilon) = \exp(-t_i \epsilon^{-\theta_s})$ for some strictly positive parameters t_i and θ_s. Thus, we find that

$$i(s,\mathcal{P}) = \frac{t_i V(s,q_i)^{\theta s}}{\sum_{j\in P} t_j V(s,q_j)^{\theta s}}, \tag{3.72}$$

so the random utility model is a Luce voting model from Theorem 2. Furthermore, applying Axiom 3, there exists "utility indicators" $u(q_j,s) = V(q_j,s)$,"popularity indicators" $t_j > 0$ and political responsiveness parameters θ_s for any candidate j for any vector of attributes $s \in S$ such that

$$i(s,\mathcal{P}) = \frac{t_i u(s,i)^{\theta_s}}{\sum_{j\in P} t_j u(s,j)^{\theta_s}}. \tag{3.73}$$

By Lemma 1 then, a random utility model with noise parameters distributed according to the Fréchet distributions $F_{i,s}(.)$ is a probabilistic voting model satisfying the three Axiom of the previous subsection.

Proving the other implication of the equivalence is a little more demanding. Assume that a probabilistic voting model satisfies axioms 1, 2 and 3. Take some attributes $s \in S$ and a candidate $i \in \mathcal{P}$. We know from Lemma 1 that there exists utility indicators $u(q_j,s)$,

popularity indicators $t_j > 0$ for any $j \in \mathcal{P}$ and a political responsiveness parameter θ_s such that

$$i(\mathcal{P}) = \frac{t_i u(q_i, s)^{\theta_s}}{\sum_{j \in P} t_j u(q_j, s)^{\theta_s}}.$$

Consider the choice between either candidate i with a popularity $t_i > 0$ with a representative utility $u(i, s)$ and n candidates with a popularity $t_j > 0$ yielding $u(j, s)$. Thus,

$$i(P) = \frac{t_i u(q_i, s)^{\theta_s}}{t_i u(q_i, s)^{\theta_s} + n t_j u(q_j, s)^{\theta_s}}.$$

Assume now that there exists two distributions $G_i(\epsilon)$ and $G_j(\epsilon)$ in $\mathcal{C}_d$ such that

$$i(\{i, j, \ldots, j\}) = \frac{t_i u(q_i, s)^{\theta_s}}{t_i u(q_i, s)^{\theta_s} + n t_j u(q_j, s)^{\theta_s}} = \int_0^\infty G_j(u(q_i, s)/u(q_j, s)\epsilon)^n dG_i(\epsilon). \tag{3.74}$$

On the other hand, consider a binary choice between candidate i and an alternative candidate k with $t_k > 0$ and $u(q_k, s)$.

$$i(\{i, k\}) = \frac{t_i u(q_i, s)^{\theta_s}}{t_i u(q_i, s)^{\theta_s} + t_k u(q_k, s)^{\theta_s}} \tag{3.75}$$

Assume that there exists some distribution $G_k(\epsilon) \in \mathcal{C}_d$ such that

$$i(\{i, k\}) = \frac{t_i u(q_i, s)^{\theta_s}}{t_i u(q_i, s)^{\theta_s} + t_k u(q_k, s)^{\theta_s}} = \int_0^\infty G_k(u(q_i, s)/u(q_k, s)\epsilon) dG_i(\epsilon). \tag{3.76}$$

Assume finally that $n t_j u(q_j, s)^{\theta_s} = t_k u(q_k, s)^{\theta_s}$ so that

$$i(\{i, j, \ldots, j\}) = i(\{i, k\}). \tag{3.77}$$

Thus, for any set of attributes $s \in S$, any platform $q_i \in Q_i$, $q_j \in Q_j$ and $q_k \in Q_k$,

$$\int_0^\infty G_k(u(q_i, s)/u(q_k, s)\epsilon) dG_i(\epsilon) - \int_0^\infty G_j(u(q_i, s)/u(q_j, s)\epsilon)^n dG_i(\epsilon) = 0. \tag{3.78}$$

As the distributions belong to the class $\mathcal{C}_d$, we can rewrite the preceding expression as

$$\int_0^\infty [G_i(\alpha_{ki}u(q_i,s)/u(q_k,s)\epsilon) - G_i(\alpha_{ji}u(q_i,s)/u(q_j,s)\epsilon)^n]dG_i(\epsilon) = 0. \tag{3.79}$$

The integrand in (3.79) must then be equal to zero for a non-zero value of ϵ. This means that for any set of attributes $s \in S$ and any platform $q_i \in Q_i$, $q_j \in Q_j$ and $q_k \in Q_k$, there exist some ϵ such that

$$G_i(\alpha_{ki}u(q_i,s)/u(q_k,s)\epsilon) = G_i(\alpha_{ji}u(q_i,s)/u(q_j,s)\epsilon)^n, \tag{3.80}$$

As $nt_j u(q_j,s)^{\theta_s} = t_k u(q_k,s)^{\theta_s}$, we obtain

$$G_i(\alpha_{ki}(\frac{t_k}{nt_j})^{1/\theta_s}\frac{u(q_i,s)\epsilon}{u(q_k,s)}) = G_i(\frac{\alpha_{ji}u(q_i,s)\epsilon}{u(q_k,s)})^n, \tag{3.81}$$

which must hold for any vector of attributes $s \in S$ and any platform $q_i \in Q_i$, $q_j \in Q_j$. We now take q_i such that $\alpha_{ji}\epsilon u(q_i,s)/u(q_j,s) = 1$. Then the preceding equation implies

$$G_i(\frac{\alpha_{ki}}{\alpha_{ji}}(\frac{t_k}{nt_j})^{1/\theta_s}) = G_i(1)^n. \tag{3.82}$$

As this must be true for any $i,j,k \in P$ and any $s \in S$, it must be that the positive constants α_{ki} and α_{ji} are such that

$$\frac{\alpha_{ki}}{\alpha_{ji}}(\frac{t_k}{t_j})^{1/\theta_s} = \alpha_0 \tag{3.83}$$

for some constant α_0 independent from i, j, k and s. Thus, (3.82) rewrites

$$G_i(\alpha_0 n^{-1/\theta_s}) = G_i(1)^n. \tag{3.84}$$

Since $G_i(1) < 1$, there exists some constant $r_i > 0$ such that

$$G_i(1) = exp(-r_i). \tag{3.85}$$

Thus,

$$G_i(\alpha_0 n^{-1/\theta_s}) = e^{-nr_i}. \tag{3.86}$$

Let $\gamma = \alpha_0 n^{-1/\theta_s}$. Injecting γ in the last equation implies

$$G_i(\gamma) = exp(-r_i(\gamma/\alpha_0)^{-\theta_s}). \tag{3.87}$$

From that point, it is straightforward that

$$\alpha_{i,j} = (r_j/r_i)^{1/\theta_s} \tag{3.88}$$

Thus, injecting (3.88) in (3.83), we find that

$$\alpha_0 = 1, \tag{3.89}$$

which proves that $G_i(.)$ follows a Fréchet distribution,

$$G_i(\gamma) = exp(-r_i(\gamma)^{-\theta_s}). \tag{3.90}$$

3.6.3 Proof of Proposition 1

Démonstration. The first-order condition writes

$$-n_s G_i(q_i, q^*_{-i}) + y\chi_i \left[\frac{\partial G_i}{\partial vs_i} \frac{\partial vs_i}{\partial \tau_{s,i}} + \sum_{j \neq i} \frac{\partial G_i}{\partial vs_j} \frac{\partial vs_j}{\partial \tau_{s,i}} \right] = 0 \tag{3.91}$$

with

$$\frac{\partial G_i}{\partial vs_i} = \frac{K \left[vs_i\right]^K \left[\sum_{j \neq i} \left[vs_j\right]^K\right]}{\left[\sum_{j \in \mathcal{P}} \left[vs_j\right]^K\right]^2} \frac{1}{vs_i} = \frac{K}{vs_i} G_i \left(1 - G_i\right) > 0 \tag{3.92}$$

$$\frac{\partial G_i}{\partial vs_j} = -\frac{K \left[vs_i\right]^K \left[vs_j\right]^K}{\left[\sum_{j \in \mathcal{P}} \left[vs_j\right]^K\right]^2} \frac{1}{vs_j} = -\frac{K}{vs_j} G_i G_j < 0 \tag{3.93}$$

and

$$\frac{\partial vs_i}{\partial \tau_{s,i}} = (1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s^i} vs_{i,s}(1-vs_{i,s}) \tag{3.94}$$

$$\frac{\partial vs_j}{\partial \tau_{s,i}} = -(1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s^i}\frac{u(c_s(q_j))^{\theta_s}u(c_s(q_i))^{\theta_s}}{\left[\sum_{l\in P} u(c_s(q_l))^{\theta_s}\right]^2} = -(1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s^i} vs_{i,s}vs_{j,s} \tag{3.95}$$

This finally writes as:

$$-n_s G_i + y\chi_i\left[\frac{\partial G_i}{\partial vs_i}\frac{\partial vs_i}{\partial \tau_{s,i}} + \sum_{j\neq i}\frac{\partial G_i}{\partial vs_j}\frac{\partial vs_j}{\partial \tau_{s,i}}\right] = 0 \tag{3.96}$$

$$-n_s G_i + y\chi_i\left[\frac{K}{vs_i}G_i\left(1-G_i\right)\frac{\partial vs_i}{\partial \tau_{s,i}} - K\sum_{j\neq i} G_i G_j \frac{K}{vs_j}\frac{\partial vs_j}{\partial \tau_{s,i}}\right] = 0 \tag{3.97}$$

Manipulation of the first order condition provides for $G_i > 0$ the conditions for each interest group s

$$-n_s + y\chi_i K(1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s^i}\left[\left(1-G_i\right)\frac{vs_{i,s}(1-vs_{i,s})}{vs_i} + \sum_{j\neq i} G_j \frac{vs_{j,s}vs_{i,s}}{vs_j}\right] = 0$$

$$\frac{y}{n}\chi_i K(1-\epsilon)\frac{\theta_s}{c_s^i}\left(1-G_i\right) = 1$$

Thus in a symmetric equilibrium for any party i, and any $k, p \in S$

$$\frac{\theta_s}{c_s^i} = \frac{\theta_p}{c_p^i} = \frac{1}{\frac{y}{n}\chi_i K(1-\epsilon)\left(1-G_i\right)}. \tag{3.98}$$

Using the budget constraint (3.22), we deduce the consumption of the agents in group s as a function of the rents extracted,

$$c_s^i = \frac{\theta_s}{\overline{\theta}}\overline{y}(1-\chi_i). \tag{3.99}$$

□

3.6.4 Proof of Proposition 2

By substituting (3.34) in (3.35), it is easy to establish that the optimal number of parties is given by

$$P^* = \frac{1}{1 + K\overline{\theta}(1-\epsilon)}[K\overline{\theta}(1-\epsilon) + \frac{y}{c}]. \tag{3.100}$$

From this point, the effects of c, K, ϵ $\overline{\theta}$, y on P^* are straightforward.

Relative to the second point of the proposition, observe y and $\overline{\theta}$ both depend on n_s, for any $s \in S$. Furthermore, $d\overline{\theta}/dn_s = (\theta_s - \overline{\theta})/n$, so $d\overline{\theta}/dn_s > 0$ when $\theta_s > \overline{\theta}$ and $d\overline{\theta}/dn_s \leq 0$ otherwise. From this point, since P^* decreases with $\overline{\theta}$, it is direct that when $\theta_s < \overline{\theta}$, then P^* increases with n_s because (i) the average reactivity of the citizenry becomes lower and (ii) the tax base y increases. By contrast when $\theta_s \geq \overline{\theta}$, then the variations of P^* with n_s are ambiguous since on the one hand the average reactivity of the citizenry becomes larger, which creates a downward pressure on P^* while on the other hand the aggregate income of the citizenry y still becomes larger, which creates an upward pressure on P^*.

3.6.5 Proof of Proposition 4

By differentiating (3.41) and (3.42) with respect to χ_l and χ_h respectively, we can show that there is a complementarity between the rents extracted by the low and the high reputation parties. Thus, considering the first-order conditions separately is not sufficient to prove the unicity of the Nash equilibrium here. We need to establish a third relationship between χ_l and χ_h by combining the two FOCs. By substituting $Z(\chi_h, \chi_l)$ from (3.41) in (3.42), we find that

$$\chi_l = \frac{\chi_h(P-1)}{\chi_h((1+K\theta)(P-1)+1) - 1}. \tag{3.101}$$

Notice that χ_l decreases with χ_h as long as the denominator is positive, which establishes the unicity of the intersection of the two loci.

We have represented the effect of an increase in t_h on the position of the equilibrium. Given that $\chi_l(\chi_h)$ is independent from the popularity parameters t_l and t_h in this third

relationship, it is direct that $\chi_l^*(t_l, t_h, P)$ decreases with t_h, while $\chi_h^*(t_l, t_h, P)$ increases with t_h. The reasoning for the effect of an increase in t_l is similar.

As $\chi_h^*(t_l, t_h, P) = \chi^S(P) = \chi_l^*(t_l, t_h, P)$ when $t_h = t_l$, from the previous comparative statics, it is clear that $\chi_h^*(t_l, t_h, P) < \chi^S(P) < \chi_l^*(t_l, t_h, P)$ when $t_l < t_h$.

Finally, when the number of low popularity parties increase, then parties decrease the level of rent they extract (notice that the complementarity between χ_l and χ_h does not create an ambiguity here). This implies that $\chi_l^*(t_l, t_h, P)$ is a decreasing function of P. Furthermore, (3.41) can be rewritten as

$$1 - \frac{\chi_l}{1 - \chi_l} K\theta(1 - vs_l) = 0, \tag{3.102}$$

with vs_l the vote share of a low popularity party, so

$$vs_l(t_l, t_h, P) = 1 - \frac{1 - \chi_l^*(t_l, t_h, P)}{K\theta\chi_l^*(t_l, t_h, P)} \tag{3.103}$$

and the expected rents extracted by a low reputation party in equilibrium are equal to

$$W(t_l, t_h, P) = y\chi_l^*(t_l, t_h, P)vs_l(t_l, t_h, P). \tag{3.104}$$

Since both $\chi_l^*(t_l, t_h, P)$ and $vs_l(t_l, t_h, P)$ are positive and decreasing functions of P, the equation $W(t_l, t_h, P) = c$ admits a unique solution $P^*(t_l, t_h)$. Finally, since low popularity parties capture less rents and have a lower vote share than in a symmetric equilibrium, then $P^*(t_l, t_h)$ is lower than the number of entrants in a symmetric equilibrium.

3.6.6 Proof of Proposition 5

The first-step for determining the electoral equilibrium consists in establishing the monotonicity of the two loci $\chi_h(\chi_l, t_l, t_h, P_l, P_h)$ and $\chi_l(\chi_h, t_l, t_h, P_l, P_h)$ from the first-order conditions.

By differentiating the first-order condition of a high popularity party (3.45) with respect

to χ_l, we find that

$$W(\chi_h, \chi_l)\frac{\partial \chi_h}{\partial \chi_l} + \frac{\chi_h}{1-\chi_h} K\theta \frac{t_h \theta P_l t_l}{1-\chi_h}\left(\frac{1-\chi_l}{1-\chi_h}\right)^{K\theta-1} \frac{1}{(P_l t_l Z + P_h t_h)^2} = 0, \tag{3.105}$$

with $W(\chi_h, \chi_l)$ the differential of the LHS of (3.45) with respect to χ_h, which is negative (note that $W(\chi_h, \chi_l)$ is not the second-order condition, since we derive with respect to χ_h, not with respect to the rents extracted by a single high reputation party).

Since the second term in the LHS of (3.105) is positive, it is direct that $\partial \chi_h / \partial \chi_l > 0$, meaning that along the equilibrium path, there is a complementarity between the rents extracted by the low and high reputation parties. Similarly, we can establish with the first-order condition of a low popularity party (3.46) that $\partial \chi_l / \partial \chi_h > 0$ as well. Thus, the unicity of the electoral equilibrium is not straightforward from the first-order conditions considered separately.

The second step consists in combining the two first-order conditions in order to establish a third relationship that in turn proves the unicity of the electoral equilibrium. To this aim, we will express Z from (3.45) and substitute its expression in (3.46). From (3.45),

$$\frac{1-\chi_h}{K\theta\chi_h} = 1 - \frac{t_h^K}{P_l t_l^K Z + P_h t_h^K}, \tag{3.106}$$

so

$$P_l t_l^K Z + P_h t_h^K = \frac{t_h^K \theta \chi_h}{\theta\chi_h - (1-\chi_h)} \tag{3.107}$$

and we deduce that

$$P_l t_l^K Z = \frac{-t_h^K \theta K \chi_h (P-h-1) + P_h t_h^K (1-\chi_h)}{\theta K \chi_h - (1-\chi_h)}, \tag{3.108}$$

As (3.46) rewrites

$$1 = \frac{\chi_l}{1-\chi_l}\theta K\left[1 - \frac{Z t_l^K}{P_l t_l^K Z + P_h t_h^K}\right], \tag{3.109}$$

we deduce that

$$1 = \frac{\chi_l}{1-\chi_l}\theta K[1 - \frac{-\theta K\chi_h(P_h-1)+P_h(1-\chi_h)}{P_l\theta K\chi_h}], \tag{3.110}$$

from which we establish that

$$\chi_l = \frac{\chi_h P_l}{\chi_h((1+\theta K)(P_l+P_h)-\theta K)-P_h}. \tag{3.111}$$

In turn, (3.111) provide a negative relationship between χ_l and χ_h, which allows to determine the unicity of the Nash equilibrium, as represented in Figure 2. Furthermore, notice that the relationship between χ_l and χ_h in (3.111) is independent from the popularity parameters. This is why it is direct that χ_l^* increases with t_l and decreases with t_h, while χ_h^* decreases with t_l and increases with t_h. This, in turn, implies that $\chi_l^*(t_l,t_h,P_l,P_h) < \chi_h^*(t_l,t_h,P_l,P_h)$.

Furthermore, by differentiating the FOCs with respect to P_l or P_h, we find that $\chi_l(\chi_h,t_l,t_h,P_l,P_h)$ (resp. $\chi_h(\chi_l,t_l,t_h,P_l,P_h)$) decreases with P_l and P_h for a given value of χ_h (resp. χ_l). This implies that when P_l or P_h increases, then the two loci intersect for strictly lower values of χ_l and χ_h. Both $\chi_l^*(t_l,t_h,P_l,P_h)$ and $\chi_h^*(t_l,t_h,P_l,P_h)$ decrease with P_l and P_h.

To prove that (3.47) admits a unique solution, consider first the equation

$$y\chi_l^*(t_l,t_h,P_l,P_h)vs_l(t_l,t_h,P_l,P_h) = c_l. \tag{3.112}$$

We know that $\chi_l^*(t_l,t_h,P_l,P_h)$ decreases with P_l and P_h. Notice that in equilibrium, the first-order condition (3.45) rewrites

$$1 - \frac{\chi_l}{1-\chi_l}\theta K(1-vs_l) = 0, \tag{3.113}$$

so we can simply express the vote share of a party with low reputation as a function of χ_l,

$$vs_l(t_l,t_h,P_l,P_h) = 1 - \frac{1-\chi_l(t_l,t_h,P_l,P_h)}{\theta K\chi_l(t_l,t_h,P_l,P_h)}. \tag{3.114}$$

From this point, it is direct that vs_l increases with χ_l, so it decreases with P_l and P_h.

Consequently, the LHS of (3.112) is a decreasing function of both P_l and P_h. This is why (3.112) admits a unique solution and defines a locus $P_l(P_h)$, which is decreasing in P_h. By the same kind of reasoning, we can establish that

$$y\chi_h^*(t_l, t_h, P_l, P_h)vs_h(t_l, t_h, P_l, P_h) = c_h. \tag{3.115}$$

admits a unique solution and defines a locus $P_h(P_l)$ that is decreasing in P_l. It is then direct that the loci $P_l(P_h)$ and $P_h(P_l)$ intersect only once.

Finally, for a given value of $P_h \geq 1$, then low reputation parties have a lower incentive to enter the race relative to a case where there is no high reputation parties because (i) they capture less rents and (ii) they get a lower vote share. Thus, the locus $P_l(P_h)$ is below the line $P_h + P_l = P_l^S$, where P_l^S is the number of low popularity parties that enter the race in a symmetric equilibrium.

By a similar token, when $P_l \geq 1$, then high reputation parties have a higher incentive to enter the race relative to a case where they only face high reputation challengers because (i) they can capture more rents and (ii) get a higher vote share. The locus $P_h(P_l)$ is above the line $P_h + P_l = P_h^S$, where P_h^S is the number of high popularity parties that enter the race in a symmetric equilibrium.

Consequently, the intersection of the two loci necessarily occurs on the subspace delimitated by the two lines $P_h + P_l = P_l^S$ and $P_h + P_l = P_h^S$, which implies that $P_h^S < P_l^* + P_h^* < P_l^S$. This concludes the proof of the third point of the proposition. The fourth point is a direct consequence of the third.

3.6.7 Proof of Proposition 6

The objective of party i rewrites

$$vs_i(\chi_i y - rP(1/2 - vs_i)) + (1 - vs_i)\frac{P/2 - 1}{P - 1}rPvs_i, \tag{3.116}$$

so

$$vs_i(\chi_i y + 1/2rP) + vs_i(1 - vs_i)\frac{-P/2}{P-1}rP. \tag{3.117}$$

The first-order equation then writes

$$-1/Pn_s + \frac{\partial vs_i}{\partial \tau_s}\{\chi_i y + 1/2rP + (1 - 2vs_i)\frac{-P/2}{P-1}rP\} = 0. \tag{3.118}$$

In a symmetric equilibrium, this simplifies to the condition given in the main text,

$$-\frac{1}{P}n_s + (1-\epsilon)\frac{n_s}{n}\frac{\theta_s}{c_s}\{\chi_i y + \frac{1}{2}r\frac{P}{P-1}\}\frac{P-1}{P^2} = 0. \tag{3.119}$$

It is then direct that $c_s/\theta_s = c_k/\theta_k$ for any pair $s, k \in S$. Following the steps of the first application, this implies that

$$c_s = \frac{\theta_s}{n\overline{\theta}}(y(1-\chi_i) - r(\frac{P}{2} - 1)). \tag{3.120}$$

Replacing θ_s/c_s in the first-order condition then, we find that

$$-\frac{1}{P} + (1-\epsilon)\overline{\theta}\frac{1}{y(1-\chi) - r(P/2-1)}\{\chi_i y + \frac{1}{2}r\frac{P}{P-1}\}\frac{P-1}{P^2} = 0, \tag{3.121}$$

from which we deduce that

$$y\chi = \frac{y - r(P/2 - 1 + (1-\epsilon)\overline{\theta}/2)}{1 + (1-\epsilon)\overline{\theta}(P-1)/P}. \tag{3.122}$$

Observe that if $r = 0$ we find the result of the first application. It is interesting to notice that in a symmetric equilibrium, the cost of forming a winning coalition is precisely equal to the expected benefit from participating to governing coalitions without being the formateur. Indeed, the expected utility of party i simplifies to

$$W = 1/Py\chi. \tag{3.123}$$

Since χ is decreasing in the number of parties, and denoting c the cost of party formation, we find that the equation $W = 1/Py\chi = c$ admits a unique solution and deduce the formula of P_c^* given in the main text.

3.6.8 Proof of Lemma 2

With the notations of the canonical model of Section 3.2, i and j are ranked first by the citizenry when

$$\min(\eta(i,P)vs_i^t, \eta(j,P)vs_j^t) > \eta(k,P)vs_k^t \text{ for any } k \in P \setminus \{i,j\}. \tag{3.124}$$

Given that

$$\eta(i,P) = \frac{\mu_i}{\sum_{k\in P}\mu_k}, \tag{3.125}$$

with μ_i distributed according to a Fréchet distribution of cdf $F(\mu) = exp(-\mu^{-K})$, we deduce i and j are ranked first when

$$\min(\mu_i vs_i^t, \mu_j vs_j^t) > \mu_k vs_k^t \text{ for any } k \in P \setminus \{i,j\}. \tag{3.126}$$

It is easy to show that the distribution of $\min(\mu_i vs_i^t, \mu_j vs_j^t)$ is given by

$$F_{ij}(\epsilon) = \exp(-vs_i^K \epsilon^{-K}) + \exp(-vs_j^K \epsilon^{-K}) - \exp(-[vs_i^K + vs_j^K]\epsilon^{-K}). \tag{3.127}$$

Indeed,

$$\Pr(\min(\mu_i vs_i^t, \mu_j vs_j^t) \geq \epsilon) = \Pr(\mu_i vs_i^t \geq \epsilon)\Pr(\mu_j vs_j^t \geq \epsilon), \tag{3.128}$$

so

$$\Pr(\min(\mu_i vs_i^t, \mu_j vs_j^t) \geq \epsilon) = 1 - \exp(-vs_i^K \epsilon^{-K} - \exp(-vs_j^K \epsilon^{-K}) + \exp(-[vs_i^K + vs_j^K]\epsilon^{-K}), \tag{3.129}$$

from which we deduce $F_{ij,s}(\epsilon)$. The probability that the pair i, j is chosen by an individual with attribute s can then be expressed as

$$ij(q_i, q_{-i})) = \int_0^{\infty} \prod_{k \in P \setminus \{i,j\}} F(\epsilon / v s_k) dF_{(ij)}(\epsilon). \tag{3.130}$$

By substituting $F_{ij}(.)$ and $F(.)$, we find the formula given in the main text.

3.6.9 Proof of Proposition 7

In a symmetric equilibrium, the first-order condition with respect to the transfers $\tau_{s,i}$ to group s simplifies to

$$-\frac{n_s}{P} + K\chi(1-\epsilon)\frac{\theta_s}{c_s}\frac{n_s}{n}[\frac{1}{2}\frac{(P-2)(2P-1)}{P^2(P-1)} + \frac{1}{2}\frac{1}{P}] = 0 \tag{3.131}$$

when the solution is interior. This implies that $\theta_s/c_s = \theta_k/c_k$ for any pair $s, k \in S$. From the budget constraint we still deduce that

$$c_s = \frac{\theta_s}{n\bar{\theta}} y(1-\chi). \tag{3.132}$$

We can substitute θ_s/c_s in the first-order condition in order to determine the optimal level of extraction χ_m in runoff elections.

3.6.10 Proof of Proposition 8

Computations leading to (44) and (45): From (43), we deduce that the aggregate consumption of the enfranchised citizens is such that

$$\sum_{s \in S_f} n_s c_s = y_f + (1-\alpha_0) y_n - (1-\alpha_0)\chi(y_n + y_f). \tag{3.133}$$

As before, the parties still set their political strategy so as to maximize their expected rents. Thus, the optimal vector of transfers still verifies

$$\theta_p/c_p = \theta_s/c_s \tag{3.134}$$

for any $p, s \in S_f$. By contrast, it is optimal to extract all the resources of the agents that are not allowed to vote, so $c_k = \alpha_0 y_k$ if $k \in S_n$. From the first-order condition, we deduce that the consumption of the agents in group $s \in S_f$ writes

$$c_s(\chi) = \frac{\theta_s}{n_f \overline{\theta}_f}(y_f + (1-\alpha_0)y_n - (1-\alpha_0)\chi(y_n + y_f)). \tag{3.135}$$

Writing the first-order condition on the optimal level of extraction χ gives

$$\chi(1-\alpha_0) = \frac{1 - \alpha_0 y_n/(y_f + y_n)}{1 + \overline{\theta}_f K(1-\epsilon)(P-1)/P}. \tag{3.136}$$

From

$$1/P\chi(y_f + y_n) = c, \tag{3.137}$$

we deduce that

$$P^* = \frac{1}{1 + \overline{\theta}_f K(1-\epsilon)}[\overline{\theta}_f K(1-\epsilon) + \frac{y_f + y_n(1-\alpha_0)}{c(1-\alpha_0)}], \tag{3.138}$$

and

$$c_s^* = \frac{\theta_s(1-\epsilon)}{n_f(1 + \overline{\theta}_f K(1-\epsilon))}(y_f + (1-\alpha_0)y_n - c(1-\alpha_0)). \tag{3.139}$$

Relative to Proposition 8, the existence of an income effect is proved in the main text. If an interest group with a political responsiveness θ_c is allowed to vote, then the consumption of the agents in group $s \in S_f$ is such that

$$c_s^* = \frac{\theta_c(1-\epsilon)}{n_{f+c}(1 + \overline{\theta}_{f+c} K(1-\epsilon))}(y_{f+c} + (1-\alpha_0)y_{n-c} - c(1-\alpha_0)). \tag{3.140}$$

with $n_{f+c} = n_f + n_c$. Thus, the relative consumption of the agent in group c is

$$[\frac{n_c c_c^*}{\sum_{s\in S_f} n_s c_s^*}]_{f+c} = \frac{n_c \theta_c}{n_f \overline{\theta}_f}. \tag{3.141}$$

By contrast when the agents in group c are not allowed to vote, their consumption is necessarily equal to $\alpha_0 y_c$ since the ruling party sets $t_c = -y_c$ and the consumption of any enfranchised interest is given by (3.139). Thus, the relative consumption of the agents in group c is

$$[\frac{\alpha_0 n_c y_c}{\sum_{s\in S_f} n_s c_s^*}]_f = \frac{\alpha_0 n_c y_c (1+\overline{\theta}_f)}{n_f \overline{\theta}_f K(y_f + (1-\alpha_0)y_n - (1-\alpha_0)c)}, \tag{3.142}$$

so

$$[\frac{n_c c_c^*}{\sum_{s\in S_f\setminus c} n_s c_s^*}]_{f+c} > [\frac{\alpha_0 n_c y_c}{\sum_{s\in S_f\setminus c} n_s c_s^*}]_f \tag{3.143}$$

is equivalent to

$$\theta_c > \frac{\alpha_0 y_c (1+\overline{\theta}_f)}{y_f + (1-\alpha_0)y_n - (1-\alpha_0)c}, \tag{3.144}$$

meaning that in relative terms, the agents in a newly enfranchised group benefit from being allowed to vote *only* when they are sufficiently responsive to public policies.

Relative to the second point of the proposition, we assume in this proof that the citizenry is sufficiently large initially so that the relative size of any group $p \in S_n$ is significantly smaller than the size of the enfranchised group n_f. Thus, we can study the extension of the franchise as an increase of $+dn_p$ of n_f that is fully compensated by a decrease of $-dn_p$ of $n - n_f$. Under this simplifying assumption, the effect of the franchise extension on the consumption of any enfranchised group $s \in S_f$ is

$$\frac{dc_s^*}{dn_p} = \frac{1}{n_f(1+\overline{\theta}_f K)}\{-(1+\theta_p)c_s^* + \alpha_0 y_p\}, \tag{3.145}$$

so it is positive when the bracketed term is positive. This establishes that when group p's responsiveness increases, then the likelihood of dc_s^*/dn_p being positive decreases (since the bracketed term decreases), meaning that any enfranchised interest group s is less likely to see

its consumption increase following the enfranchisement of interest group p. Thus, groups with a low political responsiveness are more likely to be enfranchised.

Similarly, the bracketed term above decrease with c_s^*, so it decreases with y_f, meaning that richer enfranchised interests are less likely to extend the franchise. Finally, it is increasing with α_0 as long as $c < y_n$ (meaning that the aggregate income of the disenfranchised is above the cost of party formation), so when the fiscal capacity decreases, i.e. α_0 increases, then the likelihood of dc_s^*/dn_p being positive increases, meaning that any enfranchised interest s is more likely to see its consumption increase following the enfranchisement of interest group p. This means that enfranchisement of new interest groups is more likely when the fiscal capacity is low.

Main conclusion

This dissertation presented three essays on the topics of culture and institutions. Indeed, Yet no work in the literature to our knowledge has formally studied the peculiarities of rearing practices and their effects on child development.

We were able to show that the time preferences of the parents are critical in explaining their ability to set efficient rearing strategies, since higher time horizons allow to exploit the dynamic complementarities inherent to the acquisition of cognitive/cultural capital and limit permissiveness, authoritarianism, child neglect and child maltreatment. Indeed, we show that efficient rearing strategies are neither permissive nor authoritarian but authoritative. We also demonstrated that gaps in the acquisition of cognitive/cultural capital form between income groups because poorer parents are more authoritarian. Going further on unhealthy rearing practices, we suggested that video materials and television viewing for rearing purposes can widen the gaps in the acquisition of cognitive skills across income groups. Finally, we established a relatively simple condition that predicts the occurrence and the persistence of child maltreatment.

One key influence on the nexus between child rearing practices, cognitive development and the emergence of cultural norms has been left aside in this paper, the influence of the prevailing social order. Indeed, rearing practices and cognitive representations are significantly context-dependent and can change rapidly, as it has been shown for instance in the studies of Patricia Greenfield and co-authors, e.g. Childs and Greenfield (1980); Greenfield, Maynard, and Childs (2003); Greenfield (2009) on the effect of globalization on cultural learning practices. Furthermore, the working of a given social group should carefully be assessed when designing

early intervention programs. The framework on the formation of a cultural/cognitive capital set in this paper could be a starting point for such future studies.

In the second essay of this dissertation, we have presented a theory that seeks to explain why religions have repeatedly prohibited some economic activities throughout history. We suggested that the existence of economic prohibition is linked to strategic concerns of religious leaders' internalizing the dynamic interaction between the division of labor and intergenerational cultural transmission strategies.

The theory yields two main predictions. First, during transitory periods in their evolution, economic prohibition allows religious norms to diffuse in the population because it creates a *cultural* division of labor. Economic prohibition is then strategically implemented by religious leaders eager to increase the size of their cultural group under specific conditions discussed in the main text. Second, the conditions that favor the implementation of economic prohibition are also key determinants of the existence of collusion equilibria between politics and religion. Thus, the intricate linkage between cultural evolution and occupational decisions may explain both why economic discrimination allows strong religious preferences to take root in the population and why conservative religious leaders can gain significant political leadership.

The framework of this essay could be extended so as to account for labor mobility of innovators in a fragmented polity. Indeed, as argued by Mokyr (2016), the migrations of innovators across European countries partly explain the failure of the Church to decrease the rate of scientific discoveries in the region. [15] I have left aside the role of institutions affecting occupational decisions in this paper, thereby abstracting from the rich interactions between guilds or universities, technical change and religion. [16]

Another interesting venue for future research would be to account for state dependency in technology adoption. This, ultimately, could allow for a careful analysis of the long-term

15. See as well David (2008) on the linkage between competition among Europe's noble patrons motivated to attract prestigious intellectuals and the openness of scientific knowledge.

16. Richardson and McBride (2009) suggest that guilds relied on religion to sustain cooperation when mortality rates were high. They suggest that in the sixteenth century, when the disease environment eased - and given the availability of a new religious doctrine with the Reformation - new methods or organizing industry could develop. Relatedly, de la Croix, Doepke, and Mokyr (2016) put forth the role of guilds in the development trajectory of Europe. Huff (2003) argues that the rise of universities is an important institutional turning point in the history of European science.

consequences on growth and technology adoption of past episodes of economic prohibition.[17] Finally, I have been concerned in this paper with one religious denomination. Yet a large literature has put forth the role of competition on the religious market as an important determinant of religiosity.[18] Such an extension could also help deepen our understanding of the interaction between the diffusion of the Reform and the evolution of occupational decisions in Europe.

In the third essay, we have presented a new stochastic voting model for multi-candidate elections. In the paper, we demonstrate that Fréchet (or extreme type II) distributions significantly ease the issue of computing candidates' objective functions in plurality, run-off and proportional elections when the number of candidates is arbitrary. Furthermore, by relying on an axiomatic approach to probabilistic voting theories inspired of McFadden (1974), we were able to micro-found the use of Fréchet distributions in stochastic voting models.

Our central objective in this last essay was to show that our theory is flexible and provides a unifying framework to study various topics of the political economy literature that have traditionally grown independently. Five applications of the canonical model are developed on the topics of special interest politics, coalition formation in the legislature, Duverger's law and hypothesis and franchise extension in democracy. Those applications yield various refutable predictions that are provided in the propositions. Among the key contributions of the paper, we establish that both proportional and run-off systems should lead to higher party fragmentation than a plurality system. Furthermore, we were able to show that allowing high campaign investments leads to less fragmented polities and more rent extraction from well funded political parties.

Important issues have been left aside, which could nevertheless be approached with the stochastic model presented in this paper. For instance, we have not apprehended the issue of the formation of opposition movements in autocracies. Arguably, a theory on the

17. Bezin (2016) considers state dependency in technology adoption in a model of cultural evolution.

18. See for instance Finke and Iannaccone (1993), Stark and Iannaccone (1994), Stark and Bainbridge (1996), Barros and Garoupa (2002), Montgomery (2003), McBride (2008) and Montgomery (2010). See as well the review of Iyer (2016). Relatedly, Verdier and Zenou (2015) provide an extension of their dynamic theory of cultural evolution that accounts for competition between several cultural leaders.

formation of an endogenously fragmented opposition could help researchers understand strategies of entrenched elites willing to avoid a democratic transition.[19] Furthermore, it could be particularly interesting to extend this theory in order to account for more complex technologies of party formation and study the process of selection of candidates within parties (e.g. Caillaud and Tirole (2002)). Finally, the issue of the policy motivations or ideologies (e.g. Snyder and Ting (2002)) of parties and candidates has been left aside in this paper, since we have focused on the effect of political competition on the fragmentation of the polity. Future works may account for both policy and office motivations. Such an extension could ultimately help disentangle the effect of preferences from that of pure competition on the motives of political entry under various electoral rules.

19. This issue has been studied by Acemoglu, Verdier, and Robinson (2004) and Padró i Miquel (2007) for example.

References

Acemoglu, D. (2002): "Directed Technical Change," *The Review of Economic Studies*, 69(4), 781.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2015): "Democracy, Redistribution, and Inequality," *Handbook of Income Distribution*, 2, 1885 – 1966, Handbook of Income Distribution.

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*, 115(4), 1167–1199.

Acemoglu, D., T. Verdier, and J. A. Robinson (2004): "Kleptocracy and Divide-and-Rule: A Model of Personal Rule," *Journal of the European Economic Association*, 2(2-3), 162–192.

Aidt, T. S., and P. S. Jensen (2009a): "Tax structure, size of government, and the extension of the voting franchise in Western Europe, 1860–1938," *International Tax and Public Finance*, 16(3), 362–394.

——— (2009b): "The taxman tools up: An event history study of the introduction of the personal income tax," *Journal of Public Economics*, 93(1-2), 160–175.

——— (2013): "Democratization and the size of government: evidence from the long 19th century," *Public Choice*, 157(3), 511–542.

Aimone, J. A., L. R. Iannaccone, M. D. Makowsky, and J. Rubin (2013): "Endogenous Group Formation via Unproductive Costs," *Review of Economic Studies*, 80(4), 1215–1236.

Aizer, A., and L. Stroud (2010): "Education, Knowledge and the Evolution of Disparities in Health," NBER Working Papers 15840, National Bureau of Economic Research, Inc.

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

——— (2010): *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-being*. Princeton University Press, Princeton.

Akerlof, R. J. (2016): "Value formation : the role of esteem," *Games and Economic Behavior*.

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

Algan, Y., E. Beasley, F. Vitaro, and R. E. Tremblay (2014): "The Impact of Non-Cognitive Skills Training on Academic and Non-academic Trajectories: From Childhood to Early Adulthood," Sciences po publications, Sciences Po.

Alger, I. (2010): "Public Goods Games, Altruism, and Evolution," *Journal of Public Economic Theory*, 12(4), 789–813.

Alger, I., and J. W. Weibull (2010): "Kinship, Incentives, and Evolution," *The American Economic Review*, 100(4), 1725–1758.

——— (2012): "A generalization of Hamilton's rule - Love others how much?," *Journal of Theoretical Biology*, 299, 42 – 54, Evolution of Cooperation.

——— (2013): "Homo Moralis?Preference Evolution Under Incomplete Information and Assortative Matching," *Econometrica*, 81(6), 2269–2302.

——— (2016): "Evolution and Kantian morality," *Games and Economic Behavior*, 98, 56 – 67.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011): *Personality Psychology and Economics*, vol. 4. Elsevier.

Anderson, D. R., and K. G. Hanson (2010): "From blooming, buzzing confusion to media literacy: The early development of television viewing," *Developmental Review*, 30(2), 239 – 255, Television and Toddlers: The Message, the Medium, and Their Impact on Early Cognitive Development.

Ansari, A., and R. Crosnoe (2016): "Children's hyperactivity, television viewing, and the potential for child effects," *Children and Youth Services Review*, 61, 135 – 140.

Attanasio, O. P. (2015): "The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies," *Journal of the European Economic Association*, 13(6), 949–997.

Aumann, R. J. (1959): "Acceptable points in general cooperative N-person games," in *Contribution to the theory of game IV, Annals of Mathematical Study 40*, ed. by R. D. Luce, and A. W. Tucker, pp. 287–324. University Press.

Auriol, E., and J.-P. Platteau (2017): "Religious co-option in autocracy: A theory inspired by history," *Journal of Development Economics*, pp. –.

Bae, H.-o., P. L. Solomon, and R. J. Gelles (2009): "Multiple child maltreatment recurrence relative to single recurrence and no recurrence," *Children and Youth Services Review*, 31(6), 617 – 624.

Banks, J. S., and J. Duggan (2005): "Probabilistic Voting in the Spatial Model of Elections: The Theory of Office-motivated Candidates," in *Social Choice and Strategic Decisions:*

Essays in Honor of Jeffrey S. Banks, ed. by D. Austen-Smith, and J. Duggan, pp. 15–56. Springer Berlin Heidelberg, Berlin, Heidelberg.

Baron, D. P. (1993): "Government Formation and Endogenous Parties," *The American Political Science Review*, 87(1), 34–47.

Baron, D. P., and D. Diermeier (2001): "Elections, Governments, and Parliaments in Proportional Representation Systems," *The Quarterly Journal of Economics*, 116(3), 933–967.

Baron, S. W. (1952): *A Social and Religious History of the Jews, 2nd ed. 18 vols.* Columbia University Press.

Barro, R., and R. McCleary (2003): "Religion and Economic Growth Across Countries," *American Sociological Review*.

Barro, R. J., and R. M. Mccleary (2005): "Which Countries Have State Religions?," *The Quarterly Journal of Economics*, 120(4), 1331–1370.

Barros, P. P., and N. Garoupa (2002): "An Economic Theory Of Church Strictness," *Economic Journal*, 112(481), 559–576.

Baumrind, D. (1966): "Effects of Authoritative Parental Control on Child Behavior," *Child Development*, 37(4), 887–907.

——— (1967): "Child care practices anteceding three patterns of pre-school behavior," *Genetic psychology monographs.*, 75(1), 43–88.

——— (1978): "Parental Disciplinary Patterns and Social Competence in Children," *Youth & Society*, 9(3), 239–267.

Becker, G. (1996): *Accounting for Tastes*. Harvard University Press.

Becker, G. M., M. H. Degroot, and J. Marschak (1963): "Stochastic models of choice behavior," *Behavioral Science*, 8(1), 105–142.

Becker, G. S., and K. M. Murphy (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96(4), 675–700.

Becker, G. S., and L. Rayo (2007a): "Evolutionary Efficiency and Happiness," *Journal of Political Economy*, 115, 302–337.

——— (2007b): "Habits, Peers, and Happiness: An Evolutionary Perspective," *American Economic Review*, 97(2), 487–491.

Becker, G. S., and G. J. Stigler (1977): "De Gustibus Non Est Disputandum," *The American Economic Review*, 67(2), 76–90.

Becker, S., S. Pfaff, and J. Rubin (2016): "Causes and consequences of the Protestant Reformation," *Explorations in Economic History*, 62(C), 1–25.

Becker, S. O., and L. Woessmann (2009): "Was Weber Wrong? A Human Capital Theory of Protestant Economic History," *The Quarterly Journal of Economics*, 124(2), 531–596.

Bénabou, R., D. Ticchi, and A. Vindigni (2015): "Forbidden Fruits: The Political Economy of Science, Religion, and Growth," NBER Working Papers 21105, National Bureau of Economic Research, Inc.

Bénabou, R., D. Ticchi, and A. Vindigni (2015): "Religion and Innovation," *American Economic Review*, 105(5), 346–51.

Bénabou, R., and J. Tirole (2006a): "Belief in a Just World and Redistributive Politics," *The Quarterly Journal of Economics*, 121(2), 699–746.

——— (2006b): "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652–1678.

——— (2011): "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 126(2), 805–855.

——— (2012): "Laws and norms," Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit 6290, Working Paper, Bonn.

Berger, P. (1967): *The Sacred Canopy: Elements of a Sociological Theory of Religion*, Anchor books. Doubleday.

Berman, E. (2000): "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *The Quarterly Journal of Economics*, 115(3), 905–953.

Besley, T., and R. Burgess (2002): "The Political Economy of Government Responsiveness: Theory and Evidence from India," *The Quarterly Journal of Economics*, 117(4), 1415–1451.

Besley, T., and S. Coate (1997): "An Economic Model of Representative Democracy," *The Quarterly Journal of Economics*, 112(1), 85–114.

Besley, T., and T. Persson (2013): "Taxation and Development," STICERD - Economic Organisation and Public Policy Discussion Papers Series 041, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

——— (2017): "Democratic Values and Institutions," Discussion paper, Working Paper.

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

Bezin, E. (2016): "An Economic Theory of Ethical Consumer Culture and Sustainable Technological Change," Working paper, Paris School of Economics.

Bisin, A., and T. Verdier (2000a): "A model of cultural transmission, voting and political ideology," *European Journal of Political Economy*, 16(1), 5–29.

——— (2000b): ""Beyond The Melting Pot": Cultural Transmission, Marriage, And The Evolution Of Ethnic And Religious Traits," *The Quarterly Journal of Economics*, 115(3), 955–988.

——— (2001): "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97(2), 298–319.

——— (2011): "The Economics of Cultural Transmission and Socialization," in *Social Economics Handbook*, ed. by A. B. Jess Benhabib, and M. O. Jackson, pp. 339–416. Elsevier.

——— (2015): "On the Joint Evolution of Culture and Institutions," *New York University WP*.

Black, D. (1958): *The theory of committees and elections*. University Press, Cambridge.

Blaydes, L., and E. Chaney (2013): "The Feudal Revolution and Europe's Rise: Political Divergence of the Christian West and the Muslim World before 1500 CE," *American Political Science Review*, 107(1), 16–34.

Botticini, M., and Z. Eckstein (2012): *The chosen few: How education shaped Jewish history, 70-1492*. Princeton University Press.

Boyd, R., and P. J. Richerson (1985): *Culture as the Evolutionary Process*. University of Chicago Press.

Brams, S. J., and M. K. O'Leary (1970): "An Axiomatic Model of Voting Bodies," *American Political Science Review*, 64(2), 449–470.

Buckser, A. (1996): "Religion, Science, and Secularization Theory on a Danish Island," *Journal for the Scientific Study of Religion*, 35, 432.

Caillaud, B., and J. Tirole (2002): "Parties as Political Intermediaries," *The Quarterly Journal of Economics*, 117(4), 1453–1489.

Campante, F. R., and D. H. Yanagizawa-Drott (2013): "Does Religion Affect Economic Growth and Happiness? Evidence from Ramadan," NBER Working Papers 19768, National Bureau of Economic Research, Inc.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014): "Early Childhood Investments Substantially Boost Adult Health," *Science*, 343(6178), 1478–1485.

Campbell, F. A., and C. T. Ramey (1994): "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families," *Child Development*, 65(2), 684–698.

Cantoni, D., J. Dittmar, and N. Yuchtman (2016): "Reformation and Reallocation: Religious and Secular Economic Activity in Early Modern Germany," CESifo Working Paper Series 6218, CESifo Group Munich.

Carvalho, J.-P. (2013): "Veiling," *The Quarterly Journal of Economics*, 128(1), 337–370.

Carvalho, J.-P., and M. Koyama (2016): "Jewish emancipation and schism: Economic development and religious change," *Journal of Comparative Economics*, pp. –.

Cavalli-Sforza, L. L., and M. W. Feldman (1985): *Cultural Transmission and Evolution: A Quantitative Approach.* Princeton University Press.

Certain, L. K., and R. S. Kahn (2002): "Prevalence, Correlates, and Trajectory of Television Viewing Among Infants and Toddlers," *Pediatrics*, 109(4), 634–642.

Chan, T. W., and A. Koo (2011): "Parenting Style and Youth Outcomes in the UK*," *European Sociological Review*, 27(3), 385–399.

Chaney, E. (2011): "Separation of Powers and the Medieval Roots of Institutional Divergence between Europe and the Islamic Middle East," *IEA Papers and Proceedings*.

——— (2016): "Religion and the Rise and Fall of Islamic Science," Working paper, Harvard University.

Childs, C. P., and P. M. Greenfield (1980): "Informal modes of learning and teaching: The case of Zinacanteco weaving," *Studies in cross-cultural psychology*, 2, 269–316.

Christakis, D. A., F. J. Zimmerman, D. L. DiGiuseppe, and C. A. McCarty (2004): "Early Television Exposure and Subsequent Attentional Problems in Children," *Pediatrics*, 113(4), 708–713.

Coşgel, M., and T. J. Miceli (2009): "State and religion," *Journal of Comparative Economics*, 37(3), 402–416.

Cohen, R. N., and K. A. Shepsle (1990): "Multiparty Competition, Entry, and Entry Deterrence in Spatial Models of Elections," in *Advances in the Spatial Theory of Voting*, ed. by J. M. Enelow, and M. J. Hinich. Cambridge University Press.

Conti, G., J. J. Heckman, and R. Pinto (2016): "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour," *The Economic Journal*, 126(596), F28–F65.

Coughlin, P. (1992): *Probabilistic Voting Theory*, Probabilistic Voting Theory. Cambridge University Press.

Coughlin, P., and S. Nitzan (1981): "Directional and local electoral equilibria with probabilistic voting," *Journal of Economic Theory*, 24(2), 226 – 239.

Courage, M. L., and M. L. Howe (2010): "To watch or not to watch: Infants and toddlers in a brave new electronic world," *Developmental Review*, 30(2), 101 – 115, Television and Toddlers: The Message, the Medium, and Their Impact on Early Cognitive Development.

Cunha, F., I. Elo, and J. Culhane (2013): "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation," NBER Working Papers 19144, National Bureau of Economic Research, Inc.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

Cunha, F., J. Heckman, and L. Lochner (2006): "Interpreting the Evidence on Life Cycle Skill Formation," in *Handbook of the Economics of Education, Volume 1*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 12, pp. 697–812. Elsevier, 1 edn.

Cunha, F., and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4).

Cyert, R., and M. DeGroot (1987): *Bayesian Analysis and Uncertainty in Economic Theory*, G - Reference, Information and Interdisciplinary Subjects Series. Rowman & Littlefield.

David, P. A. (2008): "The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution," *Capitalism and Society*, 3(2), 1–106.

de la Croix, D., M. Doepke, and J. Mokyr (2016): "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Pre-Industrial Economy," Working Paper 22131, National Bureau of Economic Research.

De La Croix, D., and F. Mariani (2015): "From Polygyny to Serial Monogamy: A Unified Theory of Marriage Institutions," *The Review of Economic Studies*, 82(2), 565–607.

De Roover, R. (1974): *Business, banking and economic thought in late medieval and early modern Europe : selected studies ed. by Julius Kirshner*. Kirshner, Julius Editeur.

Dekel, E., J. Ely, and O. Yilankaya (2007): "Evolution of Preferences -super-1," *Review of Economic Studies*, 74(3), 685–704.

Dickson, E. S., and K. Scheve (2010): "Social Identity, Electoral Institutions and the Number of Candidates," *British Journal of Political Science*, 40(2), 349?375.

Diermeier, D., H. Eraslan, and A. Merlo (2002): "Coalition governments and comparative constitutional design," *European Economic Review*, 46(4-5), 893–907.

Diermeier, D., and A. Merlo (2004): "An empirical investigation of coalitional bargaining procedures," *Journal of Public Economics*, 88(3-4), 783–797.

Dixit, A., and J. Londregan (1996): "The Determinants of Success of Special Interests in Redistributive Politics," *The Journal of Politics*, 58, 1132–1155.

Doepke, M., and F. Zilibotti (2014): "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission," 2014 Meeting Papers 343, Society for Economic Dynamics.

——— (2017): "Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission," *Econometrica*, 85(5), 1331–1371.

Dow, J. K., and J. W. Endersby (2004): "Multinomial probit and multinomial logit: a comparison of choice models for voting research," *Electoral Studies*, 23(1), 107 – 122.

Downs, A. (1957): "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, 65(2), 135–150.

Eaton, J., and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

Ecklund, E. H., J. Z. Park, and P. T. Veliz (2008): "Secularization and Religious Change among Elite Scientists," *Social Forces*, 86(4), 1805–1839.

Ecklund, E. H., and C. P. Scheitle (2007): "Religion among Academic Scientists: Distinctions, Disciplines, and Demographics," *Social Problems*, 54(2), 289–307.

Engerman, S. L., and K. L. Sokoloff (2000): "History Lessons: Institutions, Factors Endowments, and Paths of Development in the New World," *The Journal of Economic Perspectives*, 14(3), 217–232.

——— (2005): "The Evolution of Suffrage Institutions in the New World," *The Journal of Economic History*, 65(4), 891?921.

Engstrom, R., and R. N. Engstrom (2008): "The majority vote rule and runoff primaries in the United States," *Electoral Studies*, 27, 407–416.

Epstein, S. (1980): "The stability of behavior: II. Implications for psychological research," *American Psychologist*, 35(9), 790–806.

Esteban, J., G. Levy, and L. Mayoral (2014): "Liberty, Equality, Religiosity," Working papers, LSE.

Evans, J. H. (2011): "Epistemological and Moral Conflict Between Religion and Science," *Journal for the Scientific Study of Religion*, 50(4), 707–727.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Vardi (2004): *Reasoning About Knowledge*. The MIT Press.

Feddersen, T., I. Sened, and S. Wright (1990): "Rational Voting and Candidate Entry Under Plurality Rule," *American Journal of Political Science*, 34, 1005–1016.

Fernald, L. C., P. Kariger, M. Hidrobo, and P. J. Gertler (2012): "Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal," *Proceedings of the National Academy of Sciences*, 109(Supplement 2), 17273–17280.

Finke, R., and L. R. Iannaccone (1993): "Supply-Side Explanations for Religious Change," *The ANNALS of the American Academy of Political and Social Science*, 527(1), 27–39.

Fluke, J. D. (2008): "Child protective services rereporting and recurrence - Context and considerations regarding reseach," *Child Abuse and Neglect*, 32(8), 749 – 751.

Foster, M. E., and S. Watkins (2010): "The Value of Reanalysis: TV Viewing and Attention Problems," *Child Development*, 81(1), 368–375.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. MIT Press, Cambridge, MA, Translated into Chinesse by Renin University Press, Bejing: China.

Fujiwara, T. (2015): "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil," *Econometrica*, 83(2), 423–464.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014): "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 344(6187), 998–1001.

Giavazzi, F., and G. Tabellini (2005): "Economic and political liberalizations," *Journal of Monetary Economics*, 52(7), 1297–1330.

Giuliano, P., and N. Nunn (2016): "Understanding Cultural Persistence and Change," Discussion paper, Working Paper.

Greenfield, P. M. (2009): "Linking social change and developmental change: shifting pathways of human development.," *Developmental psychology*, 45(2), 401.

Greenfield, P. M., A. E. Maynard, and C. P. Childs (2003): "Historical change, cultural learning, and cognitive representation in Zinacantec Maya children," *Cognitive Development*, 18(4), 455 – 487, The Sociocultural Construction of Implicit Knowledge.

Greif, A. (1993): "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition," *American Economic Review*, 83(3), 525–48.

——— (2006a): "Family Structure, Institutions, and Growth: The Origins and Implications of Western Corporations," *American Economic Review*, 96(2), 308–312.

——— (2006b): *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Political Economy of Institutions and Decisions. Cambridge University Press.

Greif, A., and G. Tabellini (2010): "Cultural and Institutional Bifurcation: China and Europe Compared," *American Economic Review*, 100(2), 135–40.

——— (2017): "The clan and the corporation: Sustaining cooperation in China and Europe," *Journal of Comparative Economics*, 45(1), 1–35.

Grosfeld, I., S. O. Sakalli, and E. Zhuravskaya (2016): "Middleman Minorities and Ethnic Conflict: Evidence from Pogroms," Working paper, SSRN.

Guiso, L., P. Sapienza, and L. Zingales (2003): "People's opium? Religion and economic attitudes," *Journal of Monetary Economics*, 50(1), 225–282.

Güth, W., and M. Yaari (1992): "An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game," in *Explaining Process and Change: Approaches to Evolutionary Economics*, ed. by U. Witt, Economics, cognition, and society. University of Michigan Press.

Hair, N., J. Hanson, B. Wolfe, and S. Pollak (2015): "Association of child poverty, brain development, and academic achievement," *JAMA Pediatrics*, 169(9), 822–829.

Hanson, G., and C. Xiang (2013): "Exporting Christianity: Governance and doctrine in the globalization of US denominations," *Journal of International Economics*, 91(2), 301–320.

Heckman, J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010): "Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program," *Quantitative Economics*, 1(1), 1–46.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 2052–86.

Heckman, J. J., and C. O. Corbin (2016): "Capabilities and Skills," *Journal of Human Development and Capabilities*, 17(3), 342–359.

Heckman, J. J., and T. Kautz (2012): "Hard evidence on soft skills," *Labour Economics*, 19(4), 451–464.

——— (2013): "Fostering and Measuring Skills: Interventions That Improve Character and Cognition," Working Paper 19656, National Bureau of Economic Research.

Hinich, M. J. (1977): "Equilibrium in spatial voting: The median voter result is an artifact," *Journal of Economic Theory*, 16(2), 208 – 219.

Hinich, M. J., J. O. Ledyard, and P. C. Ordeshook (1972): "Nonvoting and the existence of equilibrium under majority rule," *Journal of Economic Theory*, 4(2), 144 – 153.

Huff, T. (2003): *The Rise of Early Modern Science: Islam, China and the West*, The Rise of Early Modern Science: Islam, China, and the West. Cambridge University Press.

Iannaccone, L., R. Stark, and R. Finke (1998): "Rationality and the "Religious Mind."," *Economic Inquiry*, 36(3), 373–389.

Iannaccone, L. R. (1992): "Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives," *Journal of Political Economy*, 100(2), 271–91.

Iannaccone, L. R., R. Stark, and R. Finke (1996): "Religion, Science, and Rationality," *American Economic Review*, 86(2), 433–437.

Inglehart, R., and C. Welzel (2010): "Changing Mass Priorities: The Link between Modernization and Democracy," *Perspectives on Politics*, 8(2), 551–567.

Iyer, S. (2016): "The New Economics of Religion," *Journal of Economic Literature*, 54(2), 395–441.

James, R. N., and D. L. Sharpe (2007): "The "Sect Effect" in Charitable Giving," *American Journal of Economics and Sociology*, 66(4), 697–726.

Jean-Philippe (forthcoming): *Islam Instrumentalized. Religion and Politics in Historical Perspective.* Cambridge University Press.

Kranton, R. E. (2016): "Identity Economics 2016: Where Do Social Distinctions and Norms Come From?," *American Economic Review*, 106(5), 405–09.

Levy, G., and R. Razin (2012): "Religious Beliefs, Religious Participation, and Cooperation," *American Economic Journal: Microeconomics*, 4(3), 121–51.

Lin, T.-M., J. M. Enelow, and H. Dorussen (1999): "Equilibrium in Multicandidate Probabilistic Spatial Voting," *Public Choice*, 98(1/2), 59–82.

Lindbeck, A., and J. Weibull (1987): "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, 52(3), 273–297.

Lindberg, D. C. (2008): *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to A.D. 1450*. University Of Chicago Press, 2 edn.

Lizzeri, A., and N. Persico (2001): "The Provision of Public Goods under Alternative Electoral Incentives," *American Economic Review*, 91(1), 225–239.

——— (2004): "Why did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's ?Age of Reform?," *The Quarterly Journal of Economics*, 119(2), 707–765.

——— (2005): "A Drawback of Electoral Competition," *Journal of the European Economic Association*, 3(6), 1318–1348.

Lizzeri, A., and M. Siniscalchi (2008): "Parental guidance and supervised learning," *The Quarterly Journal of Economics*, 123(3), 1161–1195.

Llavador, H., and R. J. Oxoby (2005): "Partisan Competition, Growth, and the Franchise*," *The Quarterly Journal of Economics*, 120(3), 1155–1189.

Luby, J. (2015): "Poverty?s most insidious damage: The developing brain," *JAMA Pediatrics*, 169(9), 810–811.

Luce, R. (1959): *Individual Choice Behavior: A Theoretical Analysis*. Wiley.

Makdisi, G. (1994): *The Rise of Colleges. Institutions of Learning in Islam and the West.* Edinburgh University Press.

Manganello, J., and C. Taylor (2009): "Television exposure as a risk factor for aggressive behavior among 3-year-old children," *Archives of Pediatrics & Adolescent Medicine*, 163(11), 1037–1045.

Martin, D. (1978): *A General Theory of Secularization*, Modern revivals in sociology. Gregg Revivals.

McBride, M. (2008): "Religious Pluralism and Religious Participation: A Game Theoretic Analysis," *American Journal of Sociology*, 114(1), 77–106.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press, New York.

Meltzer, A., and S. F. Richard (1981): "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89(5), 914–27.

Merton, R. (1970): *Science, technology & society in seventeenth century England*, Harper torchbooks. The Academy library. Harper & Row.

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

Montgomery, J. D. (2003): "A Formalization and Test of the Religious Economies Model," *American Sociological Review*, 68(5), 782–809.

——— (2010): "Intergenerational Cultural Transmission as an Evolutionary Game," *American Economic Journal: Microeconomics*, 2(4), 115–36.

Morby, J. (2002): *Dynasties of the World: A Chronological and Genealogical Handbook*, Oxford Paperbacks. Oxford University Press.

Mulligan, C. B., R. Gil, and X. S. i Martin (2004): "Do Democracies Have Different Public Policies than Nondemocracies?," *The Journal of Economic Perspectives*, 18(1), 51–74.

Myerson, R. B. (1993): "Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems," *The American Political Science Review*, 87(4), 856–869.

Myerson, R. B., and R. J. Weber (1993): "A Theory of Voting Equilibria," *The American Political Science Review*, 87(1), 102–114.

National Scientific Council on the Developing Child (2004): "Young children develop in an environment of relationships," .

——— (2005a): "Excessive stress disrupts the architecture of the developing brain," .

——— (2005b): "Persistent fear and anxiety can affect young children?s learning and development," .

——— (2012): "The Science of Neglect: The Persistent Absence of Responsive Care Disrupts the Developing Brain," .

Norris, P., and R. Inglehart (2004): *Sacred and Secular: Religion and Politics Worldwide*. Cambridge University Press, Camdridge.

North, C. M., and C. R. Gwin (2004): "Religious Freedom and the Unintended Consequences of State Religion," *Southern Economic Journal*, 71(1), 103–117.

North, D. (1981): *Structure and Change in Economic History*. Norton.

Nunn, N., and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101(7), 3221–52.

Nüssli, C. (2011): "Euratlas Project," https://www.euratlas.net/index_en.html.

Obel, C., T. B. Henriksen, S. Dalsgaard, K. M. Linnet, E. Skajaa, P. H. Thomsen, and J. Olsen (2004): "Does Children's Watching of Television Cause Attention Problems? Retesting the Hypothesis in a Danish Cohort," *Pediatrics*, 114(5), 1372–1373.

O'Connor, T. G., M. Rutter, C. Beckett, L. Keaveney, and J. M. a. Kreppner (2000): "The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-up," *Child Development*, 71(2), 376–390.

Ok, E. A., and F. Vega-Redondo (2001): "On the Evolution of Individualistic Preferences: An Incomplete Information Scenario," *Journal of Economic Theory*, 97(2), 231 – 254.

Osborne, M. (1995): "Spatial Models of Political Competition under Plurality Rule: A Survey of Some Explanations of the Number of Candidates and the Positions They Take," *Canadian Journal of Economics*, 28(2), 261–301.

Osborne, M. J. (1993): "Candidate Positioning and Entry in a Political Competition," *Games and Economic Behavior*, 5(1), 133 – 151.

Osborne, M. J., and A. Slivinski (1996): "A Model of Political Competition with Citizen-Candidates*," *The Quarterly Journal of Economics*, 111(1), 65–96.

Padró i Miquel, G. (2007): "The Control of Politicians in Divided Societies: The Politics of Fear," *The Review of Economic Studies*, 74(4), 1259–1274.

Palfrey, T. R. (1984): "Spatial Equilibrium with Entry," *The Review of Economic Studies*, 51(1), 139–156.

Persson, T., and G. Tabellini (2002): *Political Economics: Explaining Economic Policy*, Zeuthen lecture book series. MIT Press.

Piotrowski, J. T., M. A. Lapierre, and D. L. Linebarger (2013): "Investigating Correlates of Self-Regulation in Early Childhood with a Representative Sample of English-Speaking American Families," *Journal of Child and Family Studies*, 22(3), 423–436.

Platteau, J.-P. (2008): "Religion, politics, and development: Lessons from the lands of Islam," *Journal of Economic Behavior & Organization*, 68(2), 329–351.

——— (2011): "Political Instrumentalization of Islam and the Risk of Obscurantist Deadlock," *World Development*, 39(2), 243–260.

Prat, A., and D. Strömberg (2011): "The Political Economy of Mass Media," CEPR Discussion Papers 8246, C.E.P.R. Discussion Papers.

Proctor, L. J., G. A. Aarons, H. Dubowitz, D. J. English, T. Lewis, R. Thompson, J. M. Hussey, A. J. Litrownik, and S. C. Roesch (2012): "Trajectories of Maltreatment Re-Reports From Ages 4 to 12:: Evidence for Persistent Risk After Early Exposure," *Child Maltreatment*, 17(3), 207–217, PMID: 22723495.

Prummer, A., and J.-P. Siedlarek (2017): "Community leaders and the preservation of cultural traits," *Journal of Economic Theory*, 168, 143 – 176.

Putnam, R. D., R. Leonardi, and R. Nanetti (1993): *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton.

Rayo, L., and A. Robson (2013): "Biology and the Arguments of Utility," Cowles Foundation Discussion Papers 1893, Cowles Foundation for Research in Economics, Yale University.

Reed, C. G., and C. T. Bekar (2003): "Religious prohibitions against usury," *Explorations in Economic History*, 40(4), 347–368.

Richardson, G., and M. McBride (2009): "Religion, longevity, and cooperation: The case of the craft guild," *Journal of Economic Behavior & Organization*, 71(2), 172 – 186.

Rideout, V., and E. Hamel (2006): *The Media Family: Electronic Media in the Lives of Infants, Toddlers, Preschoolers and their Parents*. The Henry J. Kaiser Family Foundation.

Riker, W. H. (1982): "The Two-Party System and Duverger's Law: An Essay on the History of Political Science," *The American Political Science Review*, 76(4), 753–766.

Robalino, N., and A. Robson (2012): "The economic approach to 'theory of mind'," *Philosophical transactions of the royal society*, 367, 2224–2233.

Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg (2007): "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic

Status, and Cognitive Ability for Predicting Important Life Outcomes," *Perspectives on Psychological Science*, 2(4), 313–345, PMID: 26151971.

Robson, A. J. (1990): "Efficiency in evolutionary games: Darwin, nash and the secret handshake," *Journal of Theoretical Biology*, 144(3), 379 – 396.

——— (2001): "Why Would Nature Give Individuals Utility Functions?," *Journal of Political Economy*, 109(4), 900–929.

Robson, A. J., and L. Samuelson (2011): "Chapter 7 - The Evolutionary Foundations of Preferences," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, vol. 1, pp. 221 – 310. North-Holland.

Rowe, M. L. (2008): "Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill," *Journal of Child Language*, 35(1), 185–205.

Rubin, J. (2011): "Institutions, the Rise of Commerce and the Persistence of Laws: Interest Restrictions in Islam and Christianity," *The Economic Journal*, 121(557), 1310–1339.

Rubio-Codina, M., O. Attanasio, and S. Grantham-McGregor (2016): "Mediating pathways in the socio-economic gradient of child development," *International Journal of Behavioral Development*, 40(6), 483–491, PMID: 27885311.

Rubio-Codina, M., O. Attanasio, C. Meghir, N. Varela, and S. Grantham-McGregor (2015): "The socioeconomic gradient of child development: Cross-sectional evidence from children 6–42 months in Bogota," *Journal of Human Resources*, 50(2), 464–483.

Sagan, S. D. (1996): "Why Do States Build Nuclear Weapons?: Three Models in Search of a Bomb," *International Security*, 21(3), pp. 54–86.

Samuelson, L. (2001): "Analogies, Adaptation, and Anomalies," *Journal of Economic Theory*, 97(2), 320–366.

——— (2004): "Information-Based Relative Consumption Effects," *Econometrica*, 72(1), 93–118.

Samuelson, L., and J. M. Swinkels (2006): "Information, evolution and utility," *Theoretical Economics*, 1(1), 119–142.

Schady, N. (2011): "Parents' Educatino, Mothers' Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence From Ecuador," *American Journal of Public Health*, 101(12), 2299–2307, PMID: 22021308.

Schady, N., J. Behrman, M. C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis (2015): "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries," *Journal of Human Resources*, 50(2), 446–463.

Scherlis, G. (2014): "Political legitimacy, fragmentation and the rise of party-formation costs in contemporary Latin America," *International Political Science Review*, 35(3), 307–323.

Schofield, N. (1993): "Political competitition and multiparty coalition governments," *European Journal of Political Research*, 23(1), 1–33.

——— (1997): "Political competition," *European Journal of Political Research*, 31(1), 179–192.

——— (2007): "The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium," *The Review of Economic Studies*, 74(3), 965–980.

Schofield, N., I. Sened, and D. Nixon (1998): "Nash equilibrium in multiparty competitionwith "stochastic" voters," *Annals of Operations Research*, 84(0), 3–27.

Schwarzfuchs, S. (1966): "France and Germany under the Carolingians," in *The Dark Ages. Jews in Christian Europe, 711-1096. Editor: Cecil Roth. Revising Editor: I.H. Levine. [With Plates.].*, [World history of the Jewish people. ser. 2. vol. 2.]. London; Ramat-Gan printed.

Shepsle, K. (1991): *Models of Multiparty Electoral Competition*, Fundamentals of pure and applied economics. Harwood Academic Publishers.

Shonkoff, J. P., A. S. Garner, B. S. Siegel, M. I. Dobbins, M. F. Earls, A. S. Garner, L. McGuinn, J. Pascoe, and D. L. Wood (2011): "The Lifelong Effects of Early Childhood Adversity and Toxic Stress," *Pediatrics*, 129(1), e232–e246.

Snyder, J., and D. Strömberg (2010): "Press Coverage and Political Accountability," *Journal of Political Economy*, 118(2), 355–408.

Snyder, J. M., and M. Ting (2002): "An Informational Rationale for Political Parties," *American Journal of Political Science*, 46, 90–110.

Spratt, E., S. Pittenger, C. Swenson, A. Larosa, M. D De Bellis, M. Macias, A. P Summer, T. Hulsey, D. Runyan, and K. Brady (2012): "The Effects of Early Neglect on Cognitive, Language, and Behavioral Functioning in Childhood," *Psychology (Irvine, Calif.)*, 3, 175–182.

Stark, R. (1996): *The Rise of Christianity. A sociologist reconsiders history*. Princeton University Press & Harper, SanFrancisco.

Stark, R., and W. Bainbridge (1996): *A Theory of Religion*. Rutgers University Press.

Stark, R., and L. R. Iannaccone (1994): "A Supply-Side Reinterpretation of the "Secularization" of Europe," *Journal for the Scientific Study of Religion*, 33(3), 230–252.

Straus, M. A., and J. H. Stewart (1999): "Corporal Punishment by American Parents: National Data on Prevalence, Chronicity, Severity, and Duration, in Relation to Child and Family Characteristics," *Clinical Child and Family Psychology Review*, 2(2), 55–70.

Strömberg, D. (2004): "Radio's Impact on Public Spending*," *The Quarterly Journal of Economics*, 119(1), 189–221.

Tabellini, G. (2008): "Institutions and Culture," *Journal of the European Economic Association*, 6(2-3), 255–294.

The American Academy of Pediatrics (2001): "Children, Adolescents, and Television," *Pediatrics*, 107(2), 423–426.

Tomasello, M. (2009): *The Cultural Origins of Human Cognition*. Harvard University Press.

U.S. Department of Health and Human Services Administration on Children Youth and Families (2010): "Child Maltreatment," .

Verdier, T., and Y. Zenou (2015): "The role of cultural leaders in the transmission of preferences," *Economics Letters*, 136, 158–161.

——— (2016): "Cultural Leaders and the Dynamics of Assimilation," Discussion paper, Mimeo.

Walker, S. P., T. D. Wachs, S. Grantham-McGregor, M. M. Black, C. A. Nelson, S. L. Huffman, H. Baker-Henningham, S. M. Chang, J. D. Hamadani, B. Lozoff, J. M. M. Gardner, C. A. Powell, A. Rahman, and L. Richter (2011): "Inequality in early childhood: risk and protective factors for early child development," *The Lancet*, 378(9799), 1325 – 1338.

Weber, S. (1997): "Entry deterrence in electoral spatial competition," *Social Choice and Welfare*, 15(1), 31–56.

Weinberg, B. (2001): "An Incentive Model of the Effect of Parental Income on Children," *Journal of Political Economy*, 109(2), 266–280.

Wilson, B. (1966): *Religion in Secular Society: A Sociological Comment*, A Pelican book. Watts.

Wu, J. (2016): "Political Institutions and Preference Evolution," MPRA Paper 69597, University Library of Munich, Germany.

Zimmerman, F., and D. Christakis (2005): "Children's television viewing and cognitive outcomes: A longitudinal analysis of national data," *Archives of Pediatrics & Adolescent Medicine*, 159(7), 619–625.

Zimmerman, F., D. Christakis, and A. Meltzoff (2007a): "Television and dvd/video viewing in children younger than 2 years," *Archives of Pediatrics & Adolescent Medicine*, 161(5), 473–479.

Zimmerman, F. J., D. A. Christakis, and A. N. Meltzoff (2007b): "Associations between Media Viewing and Language Development in Children Under Age 2 Years," *The Journal of Pediatrics*, 151(4), 364 – 368.

List of figures